

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 30, 2001, 09:44:56 ; Search time 67.45 Seconds
(Without alignments)
1700.185 Million cell updates/sec

Title: US-08-846-234-5
Perfect score: 4225
Sequence: 1 MAPSKNGSNVVSFDGLND.....QVWPIDSSGSGISVIEYLF 784

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database :
- 1: SPREMBL_17:*
 - 2: sp_archaea:*
 - 3: sp_bacteria:*
 - 4: sp_fungi:*
 - 5: sp_human:*
 - 6: sp_invertebrate:*
 - 7: sp_mammal:*
 - 8: sp_mhc:*
 - 9: sp_organelle:*
 - 10: sp_phage:*
 - 11: sp_plant:*
 - 12: sp_podent:*
 - 13: sp_virus:*
 - 14: sp Vertebrate:*
 - 15: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4225	100.0	784	09ZT62	09ZT62 cucumis sat
2	2945	69.7	783	09FND9	09FND9 arabidopsis
3	1914	45.3	857	09S820	09S820 phaseolus a
4	1853	43.9	807	09SK74	09SK74 arabidopsis
5	1482	35.1	757	040077	040077 hordeum vul
6	1473	34.9	765	043408	043408 brassica ol
7	1458	34.5	773	09SCM1	09SCM1 arabidopsis
8	1413	33.5	779	09M4M7	09M4M7 persea amer
9	1140	27.0	1170	09LF27	09LF27 arabidopsis
10	697.5	16.5	386	09M442	09M442 cicer ariet
11	514.5	12.2	283	09FVM2	09FVM2 arabidopsis
12	450.5	11.6	357	09Q466	09Q466 cicer ariet
13	471.5	11.2	371	004607	004607 arabidopsis
14	350	8.3	125	042099	042099 arabidopsis
15	161	3.8	204	09AV63	09AV63 oryza sativ
16	145	3.4	88	042092	042092 arabidopsis
17	121.5	2.9	538	P74306	P74306 synechocyst
18	114	2.7	623	061391	061391 caenorhabd1
19	113.5	2.7	632	004049	004049 saccharomyc

20	111	2.6	741	10	09LNO1	09LNO1 arabidopsis
21	105.5	2.5	568	3	09P8X6	09P8X6 aspergillus
22	104	2.5	482	10	065328	065328 nicotiana g
23	103.5	2.4	1896	12	004232	004232 mucosal dis
24	103	2.4	3944	5	018667	018667 caenorhabd1
25	102.5	2.4	582	2	P73732	P73732 synechocyst
26	102.5	2.4	1093	10	09LFE6	09LFE6 arabidopsis
27	102	2.4	679	5	096176	096176 plasmodium
28	102	2.4	941	10	09LRF7	09LRF7 arabidopsis
29	101.5	2.4	626	5	061392	061392 caenorhabd1
30	101.5	2.4	626	5	09U238	09U238 caenorhabd1
31	101.5	2.4	816	2	092119	092119 lactococcus
32	101	2.4	494	11	09JLV6	09JLV6 mus musculu
33	101	2.4	612	3	09US12	09US12 schizosacch
34	101	2.4	1195	10	09C7T0	09C7T0 arabidopsis
35	100.5	2.4	564	12	067118	067118 influenza a
36	100.5	2.4	711	2	099YU3	099YU3 streptococ
37	100	2.4	448	2	09KVV7	09KVV7 streptomyce
38	99.5	2.4	2193	12	0993S1	0993S1 enterovirus
39	99.5	2.4	2193	12	098Y36	098Y36 enterovirus
40	99	2.3	474	2	09WXC1	09WXC1 thermus sp.
41	99	2.3	662	10	004372	004372 asparagus o
42	99	2.3	835	12	096E6F	096E6F potato viru
43	99	2.3	837	12	096E6F	096E6F potato viru
44	99	2.3	1002	2	09ACX6	09ACX6 vibrio chol
45	98.5	2.3	1417	10	09C9M5	09C9M5 arabidopsis

ALIGNMENTS

RESULT	ID	Query Match	Score	DB	Length	Best Local Similarity	Pred. No.	Mismatches	Indels	Gaps
1	09ZT62	100.0%	4225	10	784	100.0%	0	0	0	0
2	09ZT62	69.7%	2945	10	783	69.7%	0	0	0	0
3	09ZT62	45.3%	1914	10	857	45.3%	0	0	0	0
4	09ZT62	43.9%	1853	10	807	43.9%	0	0	0	0
5	09ZT62	35.1%	1482	10	757	35.1%	0	0	0	0
6	09ZT62	34.9%	1473	10	765	34.9%	0	0	0	0
7	09ZT62	34.5%	1458	10	773	34.5%	0	0	0	0
8	09ZT62	33.5%	1413	10	779	33.5%	0	0	0	0
9	09ZT62	27.0%	1140	10	1170	27.0%	0	0	0	0
10	09ZT62	16.5%	697.5	10	386	16.5%	0	0	0	0
11	09ZT62	12.2%	514.5	10	283	12.2%	0	0	0	0
12	09ZT62	11.6%	450.5	10	357	11.6%	0	0	0	0
13	09ZT62	11.2%	471.5	10	371	11.2%	0	0	0	0
14	09ZT62	8.3%	350	10	125	8.3%	0	0	0	0
15	09ZT62	3.8%	161	10	204	3.8%	0	0	0	0
16	09ZT62	3.4%	145	10	88	3.4%	0	0	0	0
17	09ZT62	2.9%	121.5	10	538	2.9%	0	0	0	0
18	09ZT62	2.7%	114	10	623	2.7%	0	0	0	0
19	09ZT62	2.7%	113.5	10	632	2.7%	0	0	0	0

```
Db 181 DPFALVKEAMKIVRTHLGTFRLLEKTPPGIVDKFGWCTWDAFYLTIVHPGVIEGVHLY 240
QY 241 DGGCPGGLVILIDGMOISIGHSDSPITKEGNNQYAGBOMPCLRLKPOENKFRDYVPKA 300
Db 241 DGGCPGGLVILIDGMOISIGHSDSPITKEGNNQYAGBOMPCLRLKPOENKFRDYVPKA 300
QY 301 TGPBRAGGKMAFIDELKGEFEKTEHYVNHALCGYWGGLRPVPGDPEARVIOPVLS 360
Db 301 TGPBRAGGKMAFIDELKGEFEKTEHYVNHALCGYWGGLRPVPGDPEARVIOPVLS 360
QY 361 LQMTMEDLADKTYLHKVGVLPPEKAEMEYEGLAHLEKVIDSVKIDVTHLEMLCEDY 420
Db 361 LQMTMEDLADKTYLHKVGVLPPEKAEMEYEGLAHLEKVIDSVKIDVTHLEMLCEDY 420
QY 421 GGRVDLAKAYKAMTKSINKHFKNGVYASMEHCNDPMFLGTEAISLGRVGDDEFTDPS 480
Db 421 GGRVDLAKAYKAMTKSINKHFKNGVYASMEHCNDPMFLGTEAISLGRVGDDEFTDPS 480
QY 481 GDRPGTFMLQGCCHVHCANDSLMGNFIHPDWMFQSTHPCAFAHSAISGSPITYSD 540
Db 481 GDRPGTFMLQGCCHVHCANDSLMGNFIHPDWMFQSTHPCAFAHSAISGSPITYSD 540
QY 541 SVGHNFNDLKLKLVLPDGSILRSEYALPTRDCLFEDPLHNGETMLKIMNLNKTGYIGA 600
Db 541 SVGHNFNDLKLKLVLPDGSILRSEYALPTRDCLFEDPLHNGETMLKIMNLNKTGYIGA 600
QY 601 FNCGGGMCETRRNQCFQSKRYTSKTNPKDIEHMSGENPISIEGVKTFALYLOAKK 660
Db 601 FNCGGGMCETRRNQCFQSKRYTSKTNPKDIEHMSGENPISIEGVKTFALYLOAKK 660
QY 661 LILSKPSQDLDIALDPEFELITVSPYTKLIQTSLHFAPIGLVNLMTSGAIOISVDYDD 720
Db 661 LILSKPSQDLDIALDPEFELITVSPYTKLIQTSLHFAPIGLVNLMTSGAIOISVDYDD 720
QY 721 LSSVEIGVKGCGEMRVASKRPACRIDGEGVGFKYDQDQVNVQVWPIDSSSGGIS 780
Db 721 LSSVEIGVKGCGEMRVASKRPACRIDGEGVGFKYDQDQVNVQVWPIDSSSGGIS 780
QY 781 EYLF 784
Db 781 EYLF 784

RESULT 2
Q9FND9 PRELIMINARY; PRT; 783 AA.
AC Q9FND9;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DE RAFFINOSE SYNTHASE PROTEIN.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=98069011; PubMed=9405937;
RA Kotani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. II.
RT Sequence features of the regions of 1,044,062 bp covered by thirteen
RT physically assigned pl clones."
RL DNA Res. 4:291-300(1997).
DR EMBL: A8006702; BABI1595.1; -.
SQ SEQUENCE 783 AA; 86237 MW; 3C37D1D7871888AC CRC64;
```

Query Match 69.7%; Score 2945; DB 10; Length 783;
Best Local Similarity 69.2%; Pred. No. 3.3e-222;
Matches 544; Conservative 99; Mismatches 117; Indels 26; Gaps 9;

```
QY 14 SFDGLN--DMSSPFAIDGSDFTVNGHSFLSDVPENIVASPSPTSIDKS----PVSVC 67
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 9 SDSGINGVDTEKRLDSTILANGQYVLDVYVNYLTSPYL-VKDKDVPIDVSGSF 67
QY 68 VGFDA-SEPSRHVVSIGLKDIRFMSIFPRKYWTTTHWGRNGDLESETQVILEK-- 124
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 68 IGFNLDEPKSHHVASIGLKINIRFMSIFPRKYWTTTHWGRNGDLESETQVILEK-- 127
QY 125 -----SDSGRPVYELPIYEGPRTSIOPDDDFVNVCESSGSKVYDASPRNMLLHA 178
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 128 SDSGSGSGRPVYELPIYEGPRTSIOPDDDFVNVCESSGSKVYDASPRNMLLHA 187
QY 179 GDDPFALVKEAMKIVRTHLGTFRLLEKTPPGIVDKFGWCTWDAFYLTIVHPGVIEGVH 238
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 188 GDDPFALVKEAMKIVRTHLGTFRLLEKTPPGIVDKFGWCTWDAFYLTIVHPGVIEGVH 247
QY 239 LVNDCGCPGLVILIDGMOISIGHSDSPITKEGNNQYAGBOMPCLRLKPOENKFRDYV 298
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 248 LVNDCGCPGLVILIDGMOISIGHSDSPITKEGNNQYAGBOMPCLRLKPOENKFRDYV 307
QY 299 KATGPRAGGKMAFIDELKGEFEKTEHYVNHALCGYWGGLRPVPGDPEARVIOPVLS 358
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 308 K----DQNDVGMKAFVNDLDEESTVDYIYVNHALCGYWGGLRPVPGDPEARVIOPV 363
QY 359 PGLQMTMEDLADKTYLHKVGVLPPEKAEMEYEGLAHLEKVIDSVKIDVTHLEMLCE 418
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 364 PGLQMTMEDLADKTYLHKVGVLPPEKAEMEYEGLAHLEKVIDSVKIDVTHLEMLCE 423
QY 419 DYGRVDLAKAYKAMTKSINKHFKNGVYASMEHCNDPMFLGTEAISLGRVGDDEFTD 478
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 424 KYGGRVDLAKAYKAMTKSINKHFKNGVYASMEHCNDPMFLGTEAISLGRVGDDEFTD 483
QY 479 PSGDPNCTFMLQGCCHVHCANDSLMGNFIHPDWMFQSTHPCAFAHSAISGSPITY 538
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 484 PSGDPNCTFMLQGCCHVHCANDSLMGNFIHPDWMFQSTHPCAFAHSAISGSPITY 543
QY 539 SDSVKNFNDLKLKLVLPDGSILRSEYALPTRDCLFEDPLHNGETMLKIMNLNKTGY 598
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 544 SDGVKNDLKLKLVLPDGSILRSEYALPTRDCLFEDPLHNGETMLKIMNLNKTGY 603
QY 599 GAFNCGGMCETRRNQCFQSKRYTSKTNPKDIEHMSGENPISIEGVKTFALYLOAK 658
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 604 GAFNCGGMCETRRNQCFQSKRYTSKTNPKDIEHMSGENPISIEGVKTFALYLOAK 663
QY 659 KKLILSKPSQDLDIALDPEFELITVSPYTKLIQTSLHFAPIGLVNLMTSGAIOISVD 718
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 664 KKLILSKPSQDLDIALDPEFELITVSPYTKLIQTSLHFAPIGLVNLMTSGAIOISVD 723
QY 719 DLSVEIGVKGCGEMRVASKRPACRIDGEGVGFKYDQDQVNVQVWPIDSSSGGIS 778
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 724 DE--SVEVGVGAFGEFVYASKRPVSCLDIGEVEFGY-BDSVMVQVWP---SGPGL 777
QY 779 VIEYLF 784
Db 779 SIOYLF 783

RESULT 3
Q9SBZ0 PRELIMINARY; PRT; 857 AA.
AC Q9SBZ0;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE GALACTINOL-RAFFINOSE GALACTOSYLTRANSFERASE (EC 2.4.1.67).
OS Phaseolus angulatus (Adzuki bean) (Vigna angulatus).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Vigna.
OX NCBI_TaxID=3914;
RN [1]
RP SEQUENCE FROM N.A.
```

RX MEDLINE-20117502; PubMed-10652123;
 RA Peterbauer T., Mucha J., Mayer U., Popp M., Gloessel J., Richter A.;
 RT "Synthesis of stachyose in seeds of adzuki bean (*Vigna angularis*) from
 RT Molecular cloning and functional expression of stachyose synthase from
 RT adzuki bean".
 RL Plant J. 20:509-518(1999).
 DR EMBL: Y19024; CAB64363.1; -;
 KW Transferase; Glycosyltransferase.
 SQ SEQUENCE 857 AA; 94949 MW; 85248C4B81165679 CRC64;

Query Match 45.3%; Score 1914; DB 10; Length 857;
 Best Local Similarity 43.8%; Pred. No. 2,8e-141;
 Matches 370; Conservative 139; Mismatches 243; Indels 92; Gaps 10;

17 GNDMSPPALDSDPTVNGSHSLSDVPEENIVASPSPTST---DKSPVSV----- 64
 13 GLEPSEKVFDSLSDGRLTVKGVLLSHVPEMYTF--SFSSICVPRDAPSSIIQRYTASH 70
 65 -GCFVGFPASEPDSDRHVVSIGKLDIRFMSIRFRVMTTHVGNRGDLESETQIVILE 123
 71 KGFGLGFSVSSDRLINSLSGFRGNELSTIRKTMSTOWNGSGSDLOMETQWILE 130
 124 KSDGRPVYFLPLIVEGFRITSIQPDDEFDVYCVESGSSKVVASFRSMYLHAGDDPF 183
 131 VPEY-ESYVVIPIIEKSPRSALHFGSDDHVKICAESGSTQVRASFGAIAYVVAETPY 189
 184 ALVFKAMKIVRHLSTFRLLEKTPRGIVDKFGKCTWNAFLTYHPQVIEGVRLVDOG 243
 190 NLMRAVYSLRLVHLSTFRLLEKTPRGIVDKFGKCTWNAFLTYHPQVIEGVRLVDOG 249
 244 CPGGVYLVLDGQSGHSDPTITKEGNOTVAGEOWPCALFEOENYKFRDVPNPKATGP 303
 250 VAPRVYVLDGQSGHSDPTITKEGNOTVAGEOWPCALFEOENYKFRDVPNPKATGP 309
 304 RA-----GQK-- 308
 310 NAFSPNETIKELISKIEAEHLGKQAAISAGSDLAELIMLYKVRREIDDLFGCKGK 369
 309 -----GMAFIDELKGEFKTYEHYVYVMAALGCVGGLRQVPGLEAPV 352
 370 ESNESGGCCCAACGCGMDEFTTDLRTKEFKGLDYYVMAALGCVGGLRQVPGLEAPV 428
 353 IOPVLSPLGLOMTMEDLAVDKIVLHKVGLVPRPEKAEEMEGLEHAHLEKVIDGKIVHL 412
 429 IFCKLSPGLVGMKDLAVDKIVEGSIGLVHPQANDLVSMHSTYLAQTGYTKIDVHS 488
 413 LEMLEDYGVADLAKAYYKAMTKSINKHFKNGVYASMEHCNDPMFLGTETASISLRVGD 472
 489 LEYVEEYGVGVEIAKAYYDGLTNSIIFKNGSGLIASMOQCNDFELGTQIIFGVRGD 548
 473 DWVCTDPSDDPMTWLOGCHVHNCANDSLMGNFTHPMDMFQSTHPCAAHAAASRAIS 532
 549 DWFDDPNDPMDPVGWMLQVHMHCSYNSLMGOIIPMDMFQSTHPCAAHAAASRAIS 608
 533 GGPVYVSDVSGHNDPLKLLVLPDGSILRSFYVALPTDCCLEPDDLHNGEMLKTKWNN 592
 609 GGPVYVSDVSGHNDPLKLLVLPDGSILRSFYVALPTDCCLEPDDLHNGEMLKTKWNN 668
 593 KFTGVYGAFCGCGGCRTRNOCFSQYSKRSVTSKTNPKDLEWMSGENDPISIEGVKTA 652
 669 KKGVIYGAFCGCGGCRTRNOCFSQYSKRSVTSKTNPKDLEWMSGENDPISIEGVKTA 728
 653 LYLQAKKILKSP-SQDLIALDPPEELITVSPYTKILQTSLSHAPILGLVNLMTSGA 711
 729 VYLNQAEVYHLMTPVSEPLQTLTIOPTFELVNFVPEYKLGSSINIKAPILITNMFSGGT 788
 712 IOSVYDDDLSSVEIGVKGCGEMRVASKPRACRIDGDEVGPKYDQDMVYVQWMPID 771
 789 IDLELEITE--KDYKAVKVGGRFLATSTOSPCKFOLNGSDAALFQWLPDGLTLNLAM-IE 845
 772 SSSG 775
 : |

DB 846 ENDG 849

RESULT 4 PRELIMINARY; PRT: 807 AA.
 ID 09SYJ4
 AC 09SYJ4;
 DT 01-MAY-2000 (TREMBLREL. 13, Created)
 DT 01-MAY-2000 (TREMBLREL. 13, Last sequence update)
 DT 01-OCT-2000 (TREMBLREL. 15, Last annotation update)
 DE PUTATIVE RAFINOSE SYNTHASE OR SEED IMBIBITION PROTEIN.
 GN T7B11.23 OR AT4G01970.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eustoidia II; Brassicales; Brassicaceae; Arabidopsis.
 ON NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV, COLUMBIA;
 RA Huang E.N., Nascimento L., de la Bastide M., Habermann K., Vii M.D.,
 RA Preston R.R., Spiegel L.A., See L.H., Shah R., Matero A.,
 RA O'Shaughnessy A., Rodriguez M., Shekher M., Swaby I., Schutz K.,
 RA Parnell L.D., Dedhia N.N., McCombie W.R.;
 RT "Arabidopsis thaliana BAC T7B11 from chromosome IV near 10 cM".
 RL Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Spiegel L.A., Huang E.N., Nascimento L.D., de la Bastide M., Vii D.M.,
 RA Preston R.R., Matero A., Shah R., O'Shaughnessy A., Rodriguez M.,
 RA Shekher M., Schutz K., See L.H., Swaby I., Habermann K., Dedhia N.N.,
 RA Mewes H.W., Lemcke K., Mayer K.F.X.;
 RL Submitted (Mar-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (Mar-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC007138; AAD22659.1; -;
 DR EMBL: AL161493; CAB80690.1; -;
 SQ SEQUENCE 807 AA; 90122 MW; 963DCD5A827B3388 CRC64;

Query Match 43.9%; Score 1853.5; DB 10; Length 807;
 Best Local Similarity 44.5%; Pred. No. 1.4e-136;
 Matches 366; Conservative 146; Mismatches 257; Indels 53; Gaps 14;

1 MAPSEKNGSNVSPEDGNDM--SSPFAT-----DGSDFTVNGHSLSV 43
 1 MAPLHE-----SLSSIDVIESKPLFVPITKPLIQPNSFNLSEGLCAKSDSTPLFPV 53
 44 PENIYASP-SPTSIDSKSVSV-----GCFVGFPASEPDSDRHVVSIGKLDIRFM 92
 54 PONTVFTTPSSHISISTDALPILRLVQANAHKGGFLGFKRESPODLTNSLGFEDREFL 113
 93 SIREFKVMNTTHWGNRGDLESETQIVILE--KSDGRPVYFLPLIVEGFRITSIOPE 150
 114 SIREFKVMNTTHWGNRGDLESETQIVILE--KSDGRPVYFLPLIVEGFRITSIOPE 170
 151 DDVFDVCESSGSSKVVYDASFRSMYLHAGDDPPALVYKAMKIVRHLSTFRLLEKTPPG 210
 171 KGNVLICAGSGSTKVYESSFKSIAYIHIDNPYNILKKAFAFSLRVHMFNFKLLEKTLK 230
 211 IYDKRGKCTWNAFLTYHPQVIEGVRLVNDGCGPPGLVLIDGQSGHSDLOMETQWILE 270
 231 IYDKRGKCTWNAFLTYHPQVIEGVRLVNDGCGPPGLVLIDGQSGHSDLOMETQWILE 290
 271 NOTVAGEOWPCALFEOENYKFRDVPNPKATGPRAOCCMAKFAIDELKGEFKVIDEYHVV 330
 291 NVLVGEQWTAITSTKCEKFRNKYEESLSGSDVSGSGMAAFATDLRLFRKSLDITVW 350
 331 HALCGVWGLRQVPGLEPARYIQVPLSGLOMTMEDLAVDKIVLHKVGLVPRPEKAEEM 390
 351 HALCGVWGLRQVPGLEPARYIQVPLSGLOMTMEDLAVDKIVLHKVGLVPRPEKAEEM 410

Query Match 35.1%, Score 1482; DB 10; Length 757;
Best Local Similarity 39.4%; Pred. No. 1.8e-107;
Matches 306; Conservative 128; Mismatches 266; Indels 76; Gaps 18.

391 EGLHAHLEKVGIDGKIDVILHLEMLCEDYGRVLDAAKYKAMKRSINKHFKGNGVIAS 450
Db 411 DSMHSTYASVAVGTAKIDVPTLESIAEHGGRVELAAAYIDGLTESIKKMFNGDIVIAS 470
QY 451 MEHCNDENFELTEAISTGRVGDDEFWCTDPSGDPNGTFWLOGCHMYHCANDSIWMGNFTHP 510
Db 471 MQQCNFEFFLATKQISTIGRVDDPEWMDQDPDQGVVYLQGVHMHCSYNSIWMQOMTOP 530
QY 511 DMDWFOSTHPCAAHAAASRAISGGIIVYDSVGK--NHFDLLKTKLYLDGSLIREEYAL 568
Db 531 DMDWFOSTHPCAAHAAASRAISGGIIVYDSVGK--NHFDLLKTKLYLDGSLIREEYAL 590
QY 569 PTPRCIFEDPPLHNGETMKTLINLNKFTGVIAFNGCGGMCGRFTRNOCFQSOYSKRVTSK 628
Db 591 PTPRSLFKNPLFDKRSILKIFENKFGVIGTFPNQCGAWGSPDEHRFRFGKCYTTSVSGT 650
QY 629 TNPKDIEHSGENPISIGVKTFA----LYLYQAK-LILSKPSQDLDIALDFEFELIT 683
Db 651 VHWSDIEM--DQNPFAAGSOVYTYGDYLVYKQOSEILFEMNSKSEAMKTLLEPSAFDLS 708
QY 684 VSPYTKLQTSLHFPFPIGLVNMNLNSGAIQSDYDDELSSVEIGCGGEMRFAASKRPR 743
Db 709 FVPYTELVSQGVRRAPFLDILINNFKNVGVQDMKATYGD-NSIRVDYKGGGRFMAVSSAPV 767
QY 744 ACRIDGEDVGFXYDDQD-MVYVQVFWPIDSSSGGISVIEYLE 784
Db 768 KCYLNDKEAEFEKWEETGKLSFEVFW--EEGGISHLSTFE 807

RESULT 5
Q40077 PRELIMINARY; PRT; 757 AA.
Q40077;
AC 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DF 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE SEED IMBIBITION PROTEIN.
GN STPL.
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
CC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
OC Triforceae; Hordeum.
OX NCBI_TaxID=4513;
RN [1]
RP SEQUENCE FROM N.A.
RA Heck G.R., Dorsett C., Ho T.H.;
RL Submitted (FEB-1992) to the EMBL/GenBank/DBJ databases.
DR EMBL; M7475; AAA32975.1;
DR Mendel; 16898; HORVU:1449.16898.
SQ SEQUENCE 757 AA; 82133 MW; EA9EB7B71AABBCF CRC64;

perito

2000

```

QY 255 M08IGHSS --DPIRKEGNNQVAFCEOMPCRLIKFOENCFKFDYVNPXKRTGIRAGAKGKA 312
Db 245 W0QIGSEKDDP ---GV-AV0EQAGQFASRLTGTIRENTKPF0SEHN0EET -----PGLKR 293
QY 313 FIDELKGEFFKVEHYVYHNAALCGYWGGLRPVGPB--EARYIOPVISPGL0MTMEDIAV 370
Db 294 LVDETKKE -HEVKSUYVWHAMAGYWGCVKPSAAGMEHEPALAVPV0SGVGTGN0PDITVM 352
QY 371 DKIVLHKVGLVPRPEAKEMEGEGLNAHLEKVGISDGVKIDVHLLLEMLCEDYGGRDILAAAY 430
Db 353 DLSVLGIIGLVHPRVRHREFYDELHAYILAACGVDCYKQVONIVETLIGAGHGCRVALTTRAY 412
QY 431 YKAMTKSINKHFKNGVGLASMEHCNDEMELEGTFAISLGVRVDDFMCCTP0SGDPNGTFLIQ 490
Db 413 HRALEASVARNPDPGSCSMCHNTDMLY -SAK0TAVVRASDDFPRBPAS ----- 462
QY 491 GCHWVH ---CANDSLMMONFTHPDMDFOSTHPCAAFAHSAISGDIYVSDVSGKHNF 547
Db 463 --HYHVISSVAVNLTFLGFEFQMPDMDHSLHPAEYHGAARAI0GCDIYVSDKPGNHNF 520
QY 548 DLKLKVLPGDSILRSEYALPTRDCLEFEDLNGEMLKLTWNLNFVIGAFNCG0GG 607
Db 521 DLKLKVLPGDSVYLAQPLGRPTRCLEFSDPAR0ASLKITMNMKAGVGVFC0GAG 580
QY 608 WCPREHRN0C0S0YKSRKTSKTPNDIE-----MHSGENDISIEGVKTFALYLYQ 657
Db 581 WCRVAKKTRIRIDEAPGTLVTSVRADVDALIAQAA0GTM -GGE-----AVYIAHR 629
QY 658 AKKLISKPSODLIDIALDPFEFELITVSPYTLKI0TSLHFAPIGLVNLMTSGAIO0VDY 717
Db 630 AGEIVLRLPRGATLPTLTKLREYLELHVCPL -RAVAPGVSPAPITGLHMFNAGAV0ECTV 688
QY 718 --DDDLSSVEIGVKGCGEMRVFASKRPRACRIDGEDYGVFKTID0Q -WVYVYVPP 769
Db 689 ETGEDGNAYVGLIRVAGCGRFAGYCSRRPACKSVDSDADVEFTYDSDTGLVTADVPV 744

RESULT 6
ID Q43408 PRELIMINARY; PRT; 765 AA.
AC Q43408;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, last annotation update)
DE PUTATIVE IMBIBITION PROTEIN.
OS Brassica oleracea (Cauliflower).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Brassicaceae;
OC eucosids II; Brassicales; Brassicaceae; Brassica.
CX NCBI_TaxID=3712;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-CV. ALPHA BALOMA; TISSUE=ROOT TIPS;
RA Fujikura Y., Karszen C.K.;
RA Submitted (May-1994) to the EMBL/GenBank/DBJ databases.
RL EMBL; X79330; CAA55893.1; --
DR Mendel; 30784; Braol;1449;30784.
SQ SEQUENCE 765 AA; 84084 MW; 73398603048E9B58 CRC64;

Query Match 34.9%; Score 1473.5; DB 10; Length 765;
Best Local Similarity 37.6%; Pred. No. 8.4e-107;
Matches 297; Conservative 155; Mismatches 253; Indels 85; Gaps 18;

21 MSSPAILDGSDFTVNGHSPIDVNPENIVASPSPYRIDSIPVSGCFVGFDA0SEPD0SRHV 80
Db 3 ITSNISV0NDNLVVG0KTLITKIPNIIITPVAAGSDS-----GAFIGATPK0K0SLHV 57
QY 81 VSIGKIKDIRFMSIRFKVWMTTHWVGRNGDLESETQIYLK-----SDSGRPYFL 134
Db 58 PPIGVLEGRFCRFRKLMMTQTRMGASGKDIPIETQGMLESKDEVGADAPVTVYVF 117
QY 135 LPIVGPRTTISQIPDDDFVYVCSGSSKKYVDA0FRSMYLTHADDPALVKKEMKIYR 194

```


OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Laurales; Lauraceae; Persea.
 OX NCBI_TaxID=3435;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. HASS: TISSUE-FRUIT MESOCARP;
 RA Zamorano J.P., Evans A.D., Dopico B., Lowe A.L., Wilson I.D.,
 RA Merodio C., Grierson D.;
 RT "Isolation and characterization of cDNAs for mRNAs regulated during
 RT cold storage of avocado (Persea americana Mill.) fruit."
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ13348; CAB77245.1; -
 SQ SEQUENCE 779 AA: 85368 MW: C3A8B43160316785 CRC64;

Query Match 33.5%; Score 1413.5; DB 10; Length 779;
 Best Local Similarity 36.7%; Pred. No. 4.4e-102;
 Matches 293; Conservative 144; Mismatches 268; Indels 93; Gaps 17;

QY 26 AIDGSDFTVNGHSLSDVPENIVASPSPTSIDKSPVSGCFGFDASEPDSRHVSIGK 85
 DB 8 SINDGNLVHGKTLTGVPDNIYLTPTGDL-----VAGCFIATASESISIHVFMGT 62
 QY 86 LKDIRMSIRFKVYMTTHWGRNGDLESTQIVILEKSDSGRP-----YVEL 134
 DB 63 LEGIRFCCEFFKLMWMTORMGCMGDVPLETOPLIESKDGAAAIDDEEAPTTTVE 122
 QY 135 LPYEGSPRISIOGDDDFDVCEGSSSKVVDASFSSMLTLAGDDPFALVKAAMKIVR 194
 DB 123 LPLLEGGRVAVLGDNESQIEICLESQCAVARTNOGATLVYMHAGTNPFOYINAVAYE 182
 QY 195 THLTGTFRLLEKTEPGIVDKFGMCTMDAFYLTVAPOGVEGRHLVDGCGPPGLVIDDG 254
 DB 183 KHLTSFQHLLEKKILPSFLDMGCTMDAFETDVIDEGVEELKSLSGGTPPRRLIIDGG 242
 QY 255 WOSTIGHSDPTTKGAMQYVA-GEOMPCRLKFOENKFRFYVPKATPPAGQKGMKAF 313
 DB 243 WQOIGSEETKDDSNOCVYVYEGAFASRLTGIEKENDK-----QNKSGEHPGLLV 293
 QY 314 IDELKGEEKTEHYHYMHALCGYWGGLRPOVGRP--EARVIOPLVSGLOMTMEDLAVD 371
 DB 294 VDDAK-QHHNKVEYVYHVALGTYGWGKPPAAGHEHDTALATFYVSGVGNQPDYMD 352
 QY 372 KIVLHKVGLVPERKAEMEYEGIAHLEKVGIDGVKIDVYHLEMLCEDYGGRVDLAKAY 431
 DB 353 SLVHAGLGVPRKRVFNKYNLHAKXSCGVNAYVDQNIETELGAGHGVRSLTRSYI 412
 QY 432 KAMTKSINKHFKGNGVITASMEHCDNFELGTEALISLGVGDDFMCCTDPSGDPNGTFWLOG 491
 DB 413 QALGSIARNFPDNGCIACMCHNTDSIT-SAKQRAVYVASDFFPRDAS----- 461
 QY 492 CHMYH---CANDSLWNGFIHPDMDFOSTPCAFHNASRAISGPTIYVSDVGNHND 548
 DB 462 -HTTHVSSVANNNSIFLGFQMDPDMDFSLHRAEYHQAARAVGCPITYVSDKPRHNHE 520
 QY 549 LKRLKVLDPGSLSESEYALPTRCLFEDPLHNGETMLKINLKNKFTYGICAFNGCGG 608
 DB 521 LKRLKVLDPGSLSESEYALPTRCLFEDPLHNGETMLKINLKNKFTYGICAFNGCGG 580
 QY 609 CRETRRNOCESQYKRVTSKTNPKDIE-----WISGENPISIEGVKFFALYLOAK 659
 DB 581 CKITKTRIHDAAPGTLGSIHRAHVEFINOLAGODW-NGE-----VIYFTYSGG 629
 QY 660 KLILSKPSQDDIALADPEFELITVSPYTKLIOTSLHFAPIGLVNLMTSAIQISVD--- 716
 DB 630 EVVRLPKACASIPVILEVEYELXHCIPY-KELTSWISFAPIGLIDMFSGGAVEGFVDM 688
 QY 717 -----YDDDLSS-----VEIGVKGCGEMRFAKKPRACRIDGEDV 752
 DB 689 DSNNAEPFLPGKAVKASLSSSLNNQSPSAYVVLVRGCGRGAYSSORPLKCTDLVET 748
 QY 753 GKFKDO-DQMYVYVVPMP 769

DB 749 EFNVDSVTGLVTLIIPV 766

RESULT 9
 Q9LFZ7
 ID 09LFZ7 PRELIMINARY; PRT: 1170 AA.
 AC 09LFZ7;
 DT 01-OCT-2000 (TREMUREL. 15, Created)
 DT 01-OCT-2000 (TREMUREL. 15, Last sequence update)
 DT 01-OCT-2000 (TREMUREL. 15, Last annotation update)
 DE F20N2.14.
 DE Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Khan S., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Kim C.,
 RA Shinn P., Altafi H., Bei Q., Chin C., Chlou J., Choi E., Conn L.,
 RA Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B., Lee J.,
 RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharshy N., Nguyen M.,
 RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
 RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
 RA Ecker J.R.;
 RT "Genomic sequence for Arabidopsis thaliana BMC F20N2 from chromosome
 RT 1.";
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Ecker J.R.;
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Ecker J.R.;
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Shinn P.P., Buehler E.E., Dunn P.P., Feng J.J., Kim C.C., Li Y.Y.,
 RA Walker M.M., Altafi H.H., Araujo R.R., Conn L.L., Conway A.A.B.,
 RA Gonzalez A.A., Hansen N.N.F., Huizar L.L., Kremetska I.I.,
 RA Lenz C.C., Li J.J., Liu S.S., Luros S.S., Rowley D.D., Schwartz J.J.,
 RA Toriumi M.M., Vysotskaya V.V., Yu G.G., Davis R.R.W.,
 RA Federspiel N.N.A., Theologis A.A., Ecker J.R.;
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Ecker J.R.;
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
 RA Khan S., Kim C., Altafi H., Bei B., Chin C., Chlou J., Choi E.,
 RA Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,
 RA Lee J., Lenz C., Li J., Liu A., Liu K., Liu S., Mukharshy N.,
 RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
 RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
 RA Theologis A., Theologis A., Ecker J.R.;
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC002328; AAF79504.1;
 SQ SEQUENCE 1170 AA: 129057 MW: 98B43A04E3F6D44 CRC64;

Query Match 27.0%; Score 1140; DB 10; Length 1170;
 Best Local Similarity 33.3%; Pred. No. 2.4e-80;
 Matches 259; Conservative 135; Mismatches 208; Indels 176; Gaps 20;

QY 26 AIDGSDFTVNGHSLSDVPENIVASPSPTSIDKSPVSGCFGFDASEPDSRHVSIG 84
 DB 8 SINDGNLVHGKTLTGVPDNIYLTPTGDL-----VAGCFIATASESISIHVFMGT 61
 QY 85 LKDIRMSIRFKVYMTTHWGRNGDLESTQIVILEKSDSGRPVFLFP 136

Db 62 KLEDRFMCVFRFKLMNTORNGTKNGKEIPCEIOFLIVEANQSGSDLGRRDSSSYVFLP 121
 QY 137 IVEGPEFTSIDGDDDFVDCVESGSKYVDASFRLMYLTHAGDPFALVYEAAMIV--- 193
 Db 122 IIEGFRVAVLQGENENNEIELESDDPVPDOFEGSHLVFAAGSPFVITKAKYVWFL 181
 QY 194 RTHLGT-----FRLLERKTPRGIVDKFQWCTWAFYLVHPQVIEGVRHLVD----- 241
 Db 182 KQOLKTSICLPNFRM-----PDLNMFQMGCTWDAFYNTVAKDKVQGLIESNCIDLTKPAL 235
 QY 242 -----GCCPRGLVLDGQWOSIGHDSPTKEGNNQTVAGEQPCRLKFOENYKFRD 294
 Db 236 ILCSTKAGGVTEKRYIIDDGQWVGMDETSVEFNADN-----AANFANLTHIKENHKO- 290
 QY 295 YNPKATGPRAGOKGMA-----FIDELKEFEKTEVHHVYHVALGQWGLRPOV 345
 Db 291 -----KDKRGRHVDPDLISLGHVITDLSKN-NSLKYYVNHALTYGVGKPGVS 340
 QY 346 GLP--EARYIOPVLSFGLOMTMEDLAVDKIVLHKVGLVPEPEAEEMEGVLAHLEKVGID 403
 Db 341 GMEHYSKVAVPSPSGVMSSEKCGLESITKNGLGLVPEKVFSEFYNDLHSLYASVGD 400
 QY 404 GVKIDVYHLLMECEYGGRRVDAKAYYKAMTKSTINKHFGKGVIAEMHCNDFMELGTE 463
 Db 401 GYKVDVQNTLETGLAGHGRVKLAKRYHOALEASISRNFPDNGIISCHMNTDGLY-SAK 459
 QY 464 AISLGRVGDENCTDPSGDPNGTFMLOGCHMYCAN---DSLMMGNFIHPDMDFQSTHP 520
 Db 460 KTVAVIRASDDFWRDPAS-----HTIHASVAVNTLFLGEGMQPDMDFHSLHP 508
 QY 521 CAAFHASRAISGPIYVSDSGKHNFDLKLVLDPGSLISEYALPTROCLFEDPLH 580
 Db 509 MAEYHAAARAVGCAIYVSDKPOHDFNLRLKLVLDKSGILRAKLDRGRTRELIVY----- 563
 QY 581 NGETMLKIMNLNFTVIGAFNCGGGMCEHTRNQCFSQYKRYTSKTPKDIEMHSGE 640
 Db 564 -----LPRK----- 567
 QY 641 NPISIEGVTFALYLYOAKKLILSKPSODLIALDPFEELITVSPVTKLIQTSLHPAP 700
 Db 568 -----TSLPVLMPREVEFVIVVVKERSDSS-FFAPV 599
 QY 701 GLVNLMTSGAIGSVVDLDDLSS--VEIGYKGGGEMKVFAS-KKPRACRIDGEGVSK 755
 Db 600 GLMEFNSGGAIVSLRYDDEGTRFVVMKLRGSLGVSVSRPRSVTVSDSDVEKR 657

RESULT 10
 Q9M442 PRELIMINARY: PRT: 386 AA.
 AC 09M442: (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 17, Last annotation update)
 DE PUTATIVE INHIBITION PROTEIN (FRAGMENT).
 OS Cicer arcticum (Chickpea) (Garbanzo).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC eucosids I; Fabales; Fabaceae; Papilionoideae; Cicer.
 CC NCBI_TaxID=3827;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=CV, CASTELLANA, TISSUE=ETIOLATED OSMOTIC STRESSED EPICOTYLS;
 RC Dopico B., Romo S., Labrador E.;
 RT "A putative inhibition protein is expressed in chickpea epicotyls";
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ271668; CAB71135.1; -;
 FT NON_TER 1
 SO SEQUENCE 386 AA; 42180 MW; 357AF9FBC8D71650 CRC64;

Query Match 16.5%; Score 697.5; DB 10; Length 386;
 Best Local Similarity 37.0%; Pred. No. 2.1e-46;

Matches 147; Conservative 68; Mismatches 119; Indels 63; Gaps 10;
 QY 412 LLEMLCEDYGRDLAKAYKAMTKSINKHFGKNGVIAEMHCNDEMFGTEAISGRVG 471
 Db 1 IIEFLAGHGRVSLTRSYHHALEASIRNFADNGCTIACCHNTDGLY-SAKOTAVRAS 59
 QY 472 DDFMCTDPSGDPNGTFMLOGCHMYH-----CANDSLMMGNFIHPDMDFQSTHPCAFHNAS 528
 Db 60 DDFYHPDPAS-----HTIHSSVAVNLSLFLGEFQPDWDMFHSJHPAAEYHAA 108
 QY 529 RAISGPIYVSDSGKHNFDLKLVLDPGSLISEYALPTROCLFEDPLHNGEYMLKI 588
 Db 109 RAISGPIYVSDKRGHNFDLKLVLDPGSLISEYALPTROCLFEDPLHNGEYMLKI 168
 QY 589 WNLKFTGVIGAFNCGGGMCEHTRNQCFSQYKRYTSKTPKDIEMHSGE 639
 Db 169 WNMKRCYGVGVFECQAGKCKEKKTRIDTSPGILTSSVSADVDQINQVAGVEMH-G 227
 QY 640 ENPISIEGVTFALYLYOAKKLILSKPSODLIALDPFEELITVSPVTKLIQTSLHPAP 699
 Db 228 ET-----IYAVIRSGEVIRLPKGVSLPVTILKVLFEELHFCPIOE-IAPSSIFAA 276
 QY 700 IGLVNLMTSGAIGSVVDLDDLSS--VEIGYKGGGEMKVFAS-KKPRACRIDGEGVSK 733
 Db 277 IGLDMFNTGAVEVEIHKASDKNOELFDEYVSELTLSLSPNRTKATVALKVRSGSK 336
 QY 734 MRVFSKRRACRIDGEGVSKYVDQDQ-MVVYGVPMW 769
 Db 337 FGYYSSQHPLQCAVADGIDTDPNDSGTLTFSIPV 373

RESULT 11
 Q9FVM2 PRELIMINARY: PRT: 283 AA.
 AC 09FVM2: (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 17, Last annotation update)
 DE SEED INHIBITION PROTEIN (FRAGMENT).
 OS Arabidopsis thaliana (mouse-ear cress).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
 CC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Fujiki Y., Yoshikawa Y., Sato T., Inada N., Basantli B., Ito M.,
 RT "Identification of novel dark-inducible genes from Arabidopsis
 thaliana".
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF159378; AAG23721.1; -;
 DR InterPro: IPR000719; Euk_pkinase.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
 FT NON_TER 1
 SO SEQUENCE 283 AA; 31240 MW; 33C43B84EB540301 CRC64;

Query Match 12.2%; Score 514.5; DB 10; Length 283;
 Best Local Similarity 37.2%; Pred. No. 2.9e-32;
 Matches 118; Conservative 52; Mismatches 104; Indels 43; Gaps 10;
 QY 221 DARYLVNHPGVIEGVYRHLVDSGCPGLVLIDDGQWOSIGHDSPTKEGNNQTV-AGEQM 279
 Db 1 DARYOEVTOGCVNAGLSLAGGTPPKFVILDDGQWQSVRDA-----TVEAGDEX 50
 QY 280 P-----CRLLKFOENYKFRDTPVNPKNATGPRAGOKGMAFIDELKGERKTYHHVYHVALGCG 335
 Db 51 KESPIFRLTGIKENKFK-----KKDDPNVGINKIVYKIAKEKHG-----LKYVYVNHALTG 101
 QY 336 YMGRLRQVGVGLPEARVIG-PVLSFGLOMTMEDLAVDKIVLHKVGLVPEPEAEEMEGVLAH 394

```
Db 102 YWGVR---PGEYGSVMKYPNMKSGVYENDPTWKTDMVTLQGLVSPKRYKKYNNELH 158
OY 395 AHLEKYGIDGVYIDVYIHLEMLCEDYGGVNDLAKAYKAMKRSINKHKGNGVYASMEHC 454
Db 159 SYLADAGVDGVAVVDVOCLETITLGGGLGKRVLETRKOFHALLSSVAKNPDGCKACMSHN 218
OY 455 NDPMFLGTGTAISLGRVDDFWCTDPSGDPNGTFWLOGCHMYHCAN---DSLWNGFIHPD 511
Db 219 TDAIYCSKQAAVYI-RASDFEYPRDFVS-----HTIHASVAYNSVFLGFEWQPD 266
OY 512 WDMPOSTHPCAFHNAS 528
Db 267 WDMFHSVHPAAEYHNSA 283
```

```
RESULT 12
ID 039466 PRELIMINARY; PRT: 357 AA.
AC 039466.
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE SEED INHIBITION PROTEIN (FRAGMENT).
GN SIP.
OS Cicer arietinum (Chickpea) (Garbanzo).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Cicer.
OX NCBI_TaxID=3827;
RN 11
RP SEQUENCE FROM N.A.
RA Cervantes E.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; X59875; CA65125.1; -.
DR Mendel; 7745; Cicar.1449;7745.
FT NON_TER 1
FT 357
FT 357
SQ SEQUENCE 357 AA; 39465 MW; E8CD30399E1BF3EE CRC64;
```

```
Query Match 11.6%; Score 490.5; DB 10; Length 357;
Best Local Similarity 32.5%; Pred. No. 3.2e-30;
Matches 127; Conservative 55; Mismatches 154; Indels 55; Gaps 11;

OY 187 KEAKKITYRTHIGTRLLREKTRPRGIVDKFQKCTWDATFVLVHROGVTEGVRLHVDGGRP 246
Db 3 QOAVKAAVEKNMOTGHNHREKRVPSFLDPMFGCTWDATFVLTAEVGEGLKSLSEGGR 62
OY 247 GLVLDIGWOSIGHSDPITKEGNNOTVAGEOMPCKLLK-----FOENYK--FRDYVNP 299
Db 63 RFLIM-----TWGNRLKVKQSKSGSCVVDGSAOLLGDMTRDA 101
OY 300 ATGFRAGOK-----GKAFIDELKGEFTEVHYVWHAALCGYWGSLRPOVGRPE--AR 351
Db 102 NSKKKNQONBOIPLGKHLVDGVK-KHNHYKYUWVHALLAGYWGSGVAKRATGMENHDTAL 160
OY 352 VIGPVLSPLGOMTMEDLAVDKIVLHKVGLVRRPEKAEEMTEGLAHLEKVGIDGVKIDVH 411
Db 161 ATSAVTRS--TWNPORDVIMDSLVHGLRLVHPRGFTSTN-THAYLASCGVDVDKVDVOT 217
OY 412 LLEMLCEDYGGVNDLAKAYKAMKRSINKHKGNGVYASMEHCNDPMFLGTGTAISLGRVG 471
Db 218 LIRFLVLDYVVESTLHAIIIMRLRLPLVTLNMGCIACMCHNTDGLY-SAKOTAIYVRAS 276
OY 472 DDFWCTDPSGDPNGTFWLOGCHMYH---CANDSLMGNFIHPDMFQOSTHPCAFHNAS 528
Db 277 DDFEYHPDPAS-----HTIHISVAYNSVFLGFEWQPDWDFHSLHRAEYHNSA 325
OY 529 RAISGGPIYVSDSVGKHNPDLLKKLVLPDGS 559
Db 326 RAIGGCOFMSVISOATTLILFLRLVLAOGS 356
```

```
RESULT 13
ID 004607 PRELIMINARY; PRT: 371 AA.
AC 004607;
DT 01-JUL-1997 (TREMblrel. 04, Created)
DT 01-JUL-1997 (TREMblrel. 04, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE A SIMILAR TO SEED INHIBITION PROTEIN.
GN A.IG002N01.5.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Scheet P., Magg L.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN 12
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Waterston R.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF007269; AAB61043.1; -.
DR Mendel; 16897; Arab.1449;16897.
SQ SEQUENCE 371 AA; 41564 MW; 9AEA7712D348344F CRC64;
```

```
Query Match 11.2%; Score 471.5; DB 10; Length 371;
Best Local Similarity 26.9%; Pred. No. 1e-28;
Matches 131; Conservative 71; Mismatches 112; Indels 173; Gaps 18;

OY 106 VGRNGDLESETOIVLEMS-----DSGRPVFLPIVGGPRTSIOQPDDEFVYVC 157
Db 27 MGTNGKEIFCEIOFLVEANKSGGLGDESSYVFLTEGDFRAVFOGNEANELETIC 86
OY 158 VESGSSKVVYDASFRLMLYHAAGDDPFAIVKEAMKIVRHLGTRLLREKTRPGIVDKFGW 217
Db 87 LESGK-----LQIARFEL--KLPMFDMLNMFQW 113
OY 218 CTWDATFVLV-----HQGVTEGVRLHVDGGRGLVLDIGWOSIGHSDPITKEGM 270
Db 114 CTWDATFVLVLDCCDLTKPAIILCSLK--AGVTPKFTYITDDGWSVGMDE--TSVER 167
OY 271 NOTVAGEOMPCKRLKFOENYKFRDYVNPKATGRPRAGOKGMA-----FIDELKGEF 321
Db 168 NADSA--NFANRLTHKEKHKFO-----KQKKEGHRVDDPALSLGHVITDIKSN- 215
OY 322 KTEVHYVWHAALCGYWGSLRPOVGRPEARVIOVLSFGLQNTMEDLAVDKIVLHKVGLV 381
Db 216 NSLKUYVWHAITGYWGKRPVSGI-----MSNENCGC----- 249
OY 382 PREKAEEMTEGLAHLEKVGIDGVKIDVYIHLEMLCEDYGGVNDLAKAYKAMKRSINKH 441
Db 250 -----LESTTKMGL-----GGGVKLAKKHYHDALEASISRN 279
OY 442 FKNGVYASMEHCNDPMFLGTGTAISLGRVDDFWCTDPSGDPNGTFWLOGCHMYHCAN-- 499
Db 280 FRANKRHLDLYC-----SHSQKQDLFWHHRDPAS-----HTIHASVA 317
OY 500 -DSLWNGFIHPDMFQOST--HPCAFHNASRAISGPIYVSDSVGKHNPDLLKKLV 554
Db 318 YNTFLRIGEMQPDMDISSSWRMCHLC-----QVILSDPKGQHDFFMLRKL 364
OY 555 LPDGSIL 561
Db 365 LODGSIL 371
```

```
RESULT 14
Q42099
```

ID Q42099 PRELIMINARY; PRT; 125 AA.
AC Q42099;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, last annotation update)
DE SEED IMBIBITION PROTEIN (FRAGMENT).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-WHOLE SEEDLINGS OF A THALIANA ECOTYPE COLUMBIA;
RA Desprez T., Amselem J., Chlapello H., Caboche M., Hofte H.;
RL Submitted (SEP-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; 226468; CAA81251.1; -
DR Mendel; 16899; Arath:1449;16899.
FT NON_TER 1
FT NON_TER 125
SQ SEQUENCE 125 AA; 13695 MW; 040C6E5B9BC533F CRC64;

Query Match 8.3%; Score 350; DB 10; Length 125;
Best Local Similarity 64.9%; Pred. No. 6, 2e-20;
Matches 61; Conservative 14; Mismatches 19; Indels 0; Gaps 0;

OY 522 AAFHASRAISGPIYSDSGKHNFDLKLKLVLPDGSILRSEYALPTRDCLFEDPLHN 581
DB 1 AEYHASXRAISGPIYSDSGKHNFDLKLKLVLPDGSILRRLPGRPTRDCLFADPARD 60
OY 582 GEFMLKIMNLKFTGYIGAFNCGGMCRETRRN 615
DB 61 GVSLLKIMNMNRYTGVLGYVNCGAAWSSSTERKN 94

RESULT 15
O9AV63 PRELIMINARY; PRT; 204 AA.
AC O9AV63;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE PUTATIVE SPACHYOSE SYNTHASE.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. NIPPONBARE;
RA Buell C.R., Yuan Q., Ouyang S., Moffat K.S., Hill J.N., Burr P.C.,
RA Hsiao J., Zisman V., Pal G., Bowman C.L., Fujii C.Y., VanAken S.E.,
RA Bowman C.L., Craven B., Utterback T.R., Khalak H., Feldblyum T.V.,
RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.;
RT "Oryza sativa chromosome 10 BAC OSJNB4006106 genomic sequence."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC022457; AAK27823.1; -
SQ SEQUENCE 204 AA; 21044 MW; 94605813D078387C CRC64;

Query Match 3.8%; Score 161; DB 10; Length 204;
Best Local Similarity 31.3%; Pred. No. 9e-05;
Matches 41; Conservative 23; Mismatches 49; Indels 18; Gaps 4;

OY 624 RYTSKTNRPDIEMHSGENISIEGVTFALYLQAKLLKSPSODDIALDPFERELIT 683
DB 6 KATASAPPTSMRHSGGGNGDVCFYFVEAQKQLLRNNESEFLTLEPFTEL-- 63
OY 684 VSPVTKLIQTSLHFAPIGLVNMMLNTSGAI---OSVDYDDLS---VEIGVKGCGE---- 733
DB 64 -----LLERRISFVPIGLANMLNCGAVOGFOTYVKKDDGGGDVAVAVKGAKEAYSS 117

OY 734 MRYEASKKPRA 744
DB 118 ARLMAGARGEA 128

Search completed: November 30, 2001, 09:45:02
Job time: 281 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 30, 2001, 09:41:43 ; Search time 39.46 Seconds
(without alignments)
1513.453 Million cell updates/sec

Title: US-08-846-234-5
Perfect score: 4225
Sequence: 1 MAPSFKNGSGNVSPFDGLND.....QVWPIDSSSGISVIEYLEF 784

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : PIR_68:*
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1853.5	43.9	807	2	C85025
2	1482	35.1	757	2	S27762
3	1473.5	34.9	765	2	S45033
4	1458.5	34.5	773	2	T46188
5	1140	27.0	1170	2	C96599
6	490.5	11.6	357	2	T09530
7	471.5	11.2	371	2	T01717
8	121.5	2.9	538	2	S76481
9	115.5	2.7	902	2	T41051
10	114	2.7	623	2	T42245
11	113.5	2.7	632	2	S69702
12	105	2.5	730	1	I64118
13	103	2.4	3944	2	T19957
14	102.5	2.4	564	1	HMI1V7
15	102.5	2.4	582	2	T51503
16	102.5	2.4	1093	2	T51503
17	102	2.4	679	2	G71615
18	101.5	2.4	478	2	S22622
19	101.5	2.4	626	2	T42246
20	101	2.4	514	2	A36793
21	101	2.4	612	2	T50226
22	101	2.4	1195	2	B96746
23	100.5	2.4	712	2	G02512
24	100	2.4	520	2	S27197
25	100	2.4	1550	2	S60228
26	99	2.3	770	2	S76095
27	99	2.3	1137	2	T18625
28	98	2.3	306	2	D69753
29	97.5	2.3	624	2	T42247

30	97	2.3	520	2	A25332	hydroxymethylgluta
31	97	2.3	520	2	S12736	hydroxymethylgluta
32	97	2.3	859	1	VCLJ22	env polyprotein pr
33	97	2.3	1002	2	T09438	toxR-activated lip
34	97	2.3	1013	2	B82276	toxR-activated gen
35	97	2.3	3061	1	JN0545	genome polyprotein
36	96.5	2.3	444	2	S48408	hypothetical prote
37	96.5	2.3	901	2	E84210	DNA polymerase B1
38	96.5	2.3	1095	1	A31225	phospholipase C (E
39	96	2.3	460	1	S48489	allantoinase (EC 3
40	96	2.3	520	2	S45457	hydroxymethylgluta
41	96	2.3	737	2	T24214	1,4-alpha-glucan b
42	96	2.3	1086	2	T24214	hypothetical prote
43	96	2.3	1946	2	JC6032	lactocarpin (EC 3.4
44	95.5	2.3	328	2	H83641	hypothetical prote
45	95.5	2.3	564	1	HMI1V2	hemagglutinin prec

ALIGNMENTS

RESULT 1
C85025
hypothetical protein AT4g01970 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_Change 16-Feb-2001
C:Accession: C85025
R:Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sp
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A:Reference number: A85001; MUID:20083488
A:Accession: C85025
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-807 <STO>
A:Cross-references: GB:NC_001268; NID:g7268581; PIDN:CAB80690.1; GSPDB:GN00140
A:Gene: AT4g01970
A:Map position: 4

Query Match 43.9%; Score 1853.5; DB 2; Length 807;
Best Local Similarity 44.5%; Pred. No. 8.6e+139;
Matches 366; Conservative 146; Mismatches 257; Indels 53; Gaps 14;

QY	1	MAPSFKNGSGNVSPFDGLND--SSPFAI-----DGSDFYNGHSFLSDV	43
DB	1	MAPLHE-----SLSSINDVIESKPLFVPIPKPILQPSFNLSEGSICAKDSTPLRFV	53
QY	44	PENIYASP-SPTSTIDKSPVSV-----GCFVGDASPDSPRHVVSIGKLDIRFM	92
DB	54	PONVTFPFSSHSISTDAPLPILLRVQANAHKGFGLGTRESPSRLTNSLRFDFREL	113
QY	93	SIFRRKVMWTHMGVGRNGDLESETQVILE--KSDSGRPYFLPIYEGPRTSIOPD	150
DB	114	SIFRRKVMWTHMGVGRNGDLESETQVILE--KSDSGRPYFLPIYEGPRTSIOPD	170
QY	151	DBFVDVCESSGSKKYVDSFRSMYLAHAGDDPFAVLEAMKIVRTHLGTFRLLKEKTPG	210
DB	171	KONVLICAESGSTKVESFSIAVYHICDNPYNLMKEFSAIRVHMVFKLLEKLPK	230
QY	211	IYDRGKCTWDAFYLTVPAGVIEGVRIHVDGCPPLGLVLIIDGQSGIHGDDPTTKGCM	270
DB	231	IYDRGKCTWDAFYLTVPAGVIEGVRIHVDGCPPLGLVLIIDGQSGIHGDDPTTKGCM	290
QY	271	NOTVAGEQMPRLKFOENYKFRDYVNPKATGPRAGOKMKAFIDELGEEFTVHHVVM	330
DB	291	NYLVGEGQNTALRTSFKCKKRNKKEESGSDVSGSMAFMTDLRLRFSLDDIYVM	350
QY	331	HALCGWGLRPQVGLPRAVYIOPVLSPGLOMTMEDLAVDKIVLHKYGLVPEKAEEMY	390
DB	351	HALCGAMNVRPETHMDLAKAVAPFELSPSLGATMADLAVDKVIVAGIGLVHPSRAHEPY	410

Qy	391	EGLAHLEKQIDGVKIDVYIHLLMEICEDYGGRRVDAKYYAKMKSIKNKPKGKGVAS	450
		: : : : : : : : : : : : :	
Db	411	DSMHSTASYASGVGAKIDFQGLTESIAEHHGSGVETAKYYDGLTESMIKNKNGVDIAS	470
Qy	451	MEHCNDPMFVGTETASISLGRVDDFMCTDESGDPNGTFWLQCGHMYHCANDSIWMGNFTIP	510
		: : : : : : : : : : : : : : : : :	
Db	471	MOQCNFEFFLATIKOISIGRGGDDFMWQDYPGRPOGYWYLDGYHMIKHCYNSIWMGQMTOP	530
Qy	511	DWDMFQSTHPCQAFHAASRAISGGPIYVSDSVGK - HNFDLKLKVLDPDGSLNSEYAL	568
		: : : : : : : : : : : : : :	
Db	531	DWDMFQSDHYCAEHAASRAISCGPYLLDHLGKASHNEDLTKKLAFPDGTIPRCVHYAL	590
Qy	569	PTBRCLEEDDLHNGEMTLKIMNLNKTGYIGAPNCGGCMGCFETRNOCFOSYSKRYVK	628
		: : : : : : : : : : : : : : :	
Db	591	PTBRLSKNPLFDEKSLTLKFNPKFGVIGTFNCGAGMSPEHRFKGYKCEYTYVSGT	650
Qy	629	TNPDIEMHSGENPISIEGYKTPA---LYLYQAKK-LILSPSODDLIALDPFEFEILT	683
		: : : : : : : : : : : : : : : :	
Db	651	VHVSIDIEM--DQNEPAAAGSQVYITGDYLVYKQSGSEILFLMNSKSEAMKTLEPSAFDLS	708
Qy	684	VSPVTKLIQTSLHPARIGLVNMLNTSGAIQASVDYDDLLSVELGYKCGGEMRVEFAKPR	743
		: : : : : : : : : : : : : : : :	
Db	709	FVPTETELVSSGVFEAPGLITLMNCGVQVQDMKVYTG-D-NSIVDYKGGREMAVSSAPV	767
Qy	744	ACRIDGEDVGFXYDQD-MVYVQVPMPIRDISSSGSIVIEYLX	784
		: : : : : : : : : : : : : : : : :	
Db	768	KCYLINDKEAEFKMEETGKISFFVPPV--EESGSHLSLFTF	807

```

RESULT      2
S27762
Slp1 protein - barley
C:Species: Hordeum vulgare (barley)
C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 09-Sep-1997
C:Accession: S27762
R:Heck, G.R.; Dorset, C.; Ho, T.H.
Submitted to the EMBL Data Library, February 1992
A:Description: Cloning and characterization of a gene, Slp1, associated with seed imbibition
A:Reference number: S27762
A:Accession: S27762
A:Molecule type: DNA
A:Residues: 1-757 <HEC>
A:Cross-references: EMBL:M77475; NID:g167099; PID:g167100
C:Genetics:
A:Introns: 64/2; 151/1; 181/2; 197/3; 489/3; 557/2

Query Match          35.1%   Score 1482; DB 2; Length 757;
Best Local Similarity 39.4%   Pred. No. 2.5e+109;
Matches 306; Conservative 128; Mismatches 266; Indels 76; Gaps 18;

Oy    28 DGSDFTVNGHSLSDVPRNIVASPSPYTSIDKSPYSGCFVGFDASEPDSRRHVVSIGRLK 87
      ||| :|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db    11 DGR-LAVAGRGVLVGVDPNVNTAAHAAGAIVD-----GAFVGATATAEAKSHHVFETGLR 64

Oy    88 DIRFMSIRFKWMTTHHWGRNGDLSEFTQVILE-----KDSGGPPYFL 134
      |||::||||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db    65 DCRFCMLFRFKIMMWTQRKGSTGRDVPLETQFIIEVPAAGNDGDSSDGDSSEPVYLVM 124

Oy    135 LPIVEGPRTSIOPGDDDFVDVCESGSSKYVDASFBSMLYLPHAGDDPFALVKEMAKITVR 194
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db    125 LPLLDGGQRTYLVQGNDOBELQICISGDKAVETBEGMMNNVYNHATINFPDIITQAVKAVE 184

Oy    195 THLGTRLLREKTPRGVIDKFGCWTDAPLYLTNRQVIEGVRIHLVDGCGRPGLVLIDDG 254
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db    185 KITQTFHNHRKKTVYSFDMVEGCTWDAAFYEDVTADGQKQGISLAEGGAPRFRLIIDDG 244

Oy    255 WOSIGHDS--DPTTEGMNQTVAGEOMPCKILKFQENTKFRDYVPKATGPRAGQGKMA 312
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db    245 WOQJSENKDDP----GY-AVDEGAQFASRLTGIRENKRFQSEHHQDET-----PGLKR 293

Oy    313 FIDELGGEKTYENHYVMAHALCGYWGLRPOVGPL--EARVIQPLVSLQGMTMEDIAV 370
      :|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

```

Db	294	LVDTEKKE--HGVSUYVMHMAAGYMGWGVRPSAAGMEHEEPALATYEVOSPGYVGNOPDVM	352
Qy	371	DKYLVHVGVLPEPEKAEEMVEGJLHLEKVGIDGVYIDVILHLEMLCDDGYGVDLAAAY	430
Db	353	DSLTVGLGVLGHPRHRYREYDELHATLAAAGVGVADVQNLVEYTLGAGHGGRVALLTAY	412
Qy	431	YKATKTSINKHFKNGVYASMEHCNDFMEJGTEALISLGRVGDDEFCTDPSGDPNGTFLWQ	490
Db	413	HRLEASVAAINFDPNNGCISCMCHNTDMLT-SAKQTPAVRASDPFPRDPA-----	462
Qy	491	GCHMVH--CANDSLMWKNFIHPDWMFOSTHPCAAFAAASAIISGPIYSDSYGKINF	547
Db	463	--HTVHISVAAYNTLFLGEEPMQDPMDFHSILHPAAYHGAARAIGCPIYSDSKGNINF	520
Qy	548	DLKKLYLPGSITLRSYTYALPTRDCLFEDPLNGETMLKINLKLFGYIGAFGCGGG	607
Db	521	DLKKLYLPGSVYIRALPGPRDCLFSDPARDGAASLTKINMKKCAVGVVFCQAG	580
Qy	608	WCRETRNOCFSQYRSKVTSKTNPKDIE-----HASENPDISIEGVKTFALYDQ	657
Db	581	WCRRAKKTTRIIDEAPGTLTYSVADEVALAQAAGTGM-GGE-----AVVAYHR	629
Qy	658	AKKLLSKPSODDIALDPPEFLITVSPYTKLIQTSLHPADIGLYNNLNTSGAIQSDY	717
Db	630	AGELVRLPRGATLPVTLKRLLEYELFHCYV-RAVAGVGSFADIGLHMFNAGAAECTV	688
Qy	718	---DDDLSSVEIGKGGGEMRAVASKKPRACRIDGSDVGFKKDQDQ--WVVYGVWPR	769
Db	689	ETGDGNAVYGLRVRGGRGCAVCSRRPACSVDSADVEFTYDSDTGLVTADVPV	744

Query	3	34.9%	Score 1473.5	DB 2	Length 765
Query Match					
Best Local Similarity		37.6%	Pred. No. 1.2e-108		
Matches 297		Conservative 155	Mismatches 253	Indels 85	Gaps 16
<p> RESULT 3 S45033 Probable imbibition protein - wild cabbage C:Species: Brassica oleracea (wild cabbage) C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Sep-1997 C:Accession: S45033 R:Fujikura, Y.; Karssen, C.K. Submitted to the EMBL Data Library, May 1994 A:Description: Cauliflower cDNA encoding a putative imbibition protein. A:Reference number: S45033 A:Accession: S45033 A:Status: preliminary A:Molecule type: mRNA A:Residues: 1-765 <FU> A:Cross-references: EMBL:X79330; NID:G488786; PID:G488787 </p>					
Query	21	MSSPAIDGSDFTVNGHSFLSDVDPENIVASPSYTSIDKSPVSVGCEFGDASEPDSRHV	80		
Db	3	ITSNTSVQNDNLVVGKRTILTKIPNIIITLPVAGAGSDS----	GAFIGATPFKQSKSHV	57	
QY	81	VSIGLKLDIRFSIRFKYMTTHHWGNRNGSLSEOTVILEK-----	SDSGPRVFL	134	
Db	58	PTGVLEGIRFCPCRFKLMWTQNRGASGKDIPLETQPMLESDENVAGDAPVITYVF	117		
QY	135	LPVIGPPTFSIOPDDDFVDVCEVSGSSKVVDAFSRMLYLHAGDDPFALVKEAMKIVR	194		
Db	118	LPRLGQFRAVLQNGEKNKEITCELSCKAVAGTSQGTHLVYVHAGTNPFVITQSVKAAE	177		
QY	195	THLGFRLLEKTPPGCIYDKFGSCITMDAFYLTVHQGYIEGRIHLVDGCGPPGLVLLIDG	254		
Db	178	ROMQGFHHREKKRLSPFVDMFGWCITMDAFYDVTVAEGVDEGIRLSISEGTPRPRFLIIDG	237		
QY	255	WOSIGH----DSPIRKEGNNQTVAGEOMPCRLIKFOEANKKFDYNNPKATGPRAGQKGMK	311		
Db	238	WOIENKEDSNCLVQE-----GAQFATRLVGIKENAKFKGN-DPKDT----	QVSGILK	285	
QY	312	AFIDELKGFEKTEVHYVHNLACITWGGLRQVPCPLP--EAAVIVPVLSPGLQMTMEDLA	369		


```

Db 286 SVYDNAK -ORHNVKQYAAHLAGYWGKVPRAASGMEHDSALATAYVQSPVLGNPDLY 344
Qy 370 VKIYLHKGLVLPPEAEEMYGSLAHLEKVIDGYKIDVILHLEMLCEDYGGRLAKA 429
Db 345 MOSLAVHGLVNPVKVFNFNELHSHYLASCGIDGVKVVONITETLGLGGLGRLSTRS 404
Qy 430 YKKAATKSIKHKFGKGVASMEHCNDMFELGTEATSLGRVGDDEFCPTDPSGDPNCTFYL 489
Db 405 YVQALEASTARNFKNKNGCISCHCHNTDGLY-SAKOTAIYRASDDIYPRDPAS----- 455
Qy 490 QGCHVHYCAN---DSLAMGNFTHPDMFQSTHPCAFHAASRAISGPIYVSDYVKHN 546
Db 456 ---HTHIASVAVNLTFLGEFQPDMDHSLHPTEAYHAARAAGCAIYVSDKRGNNH 512
Qy 547 FOLLKLVLPDGSILRSEYVALPTRCLEFEDPLHNGETMLKTNLKKFTGVIGAFNGCG 606
Db 513 FOLLRLVLPDGSILRSEYVALPTRCLEFEDPLHNGETMLKTNLKKFTGVIGAFNGCGA 572
Qy 607 GACRETRRQCSQVSKRYTSKTNPKDI-----EMHSGENPISIGVKTFALYLYQ 657
Db 573 GCKDKTKKRIRHDTSPGTLGLVRAEDADLISEVAGQDM-GGDS-----IYATK 621
Qy 658 AKKLLSKPSQDLIALDPFEFELTVSPVTKLIQTSLSHAPIGLVNMLNTSGAIOGV-- 715
Db 622 SEGLVRLPKGASIPITLKVLEYELFHISPL-KEITASISFAPIGLDMNSSGAIQSMEL 680
Qy 716 ----DYDDLSS-----VEIGVKGCGEMVRASKPRACRIDGEDYGVKTYQ 758
Db 681 NVTVEKPEPLSSSVSENRSPALISLGRGCRGAYASORPLCAVDGETEPNYDA 740
Qy 759 D-OMVVQVVP 767
Db 741 EVGLVTLNLP 750

RESULT 4
746188
1mb1b1on protein homolog - Arabidopsis thaliana
N:Alternate names: protein T8H10.120
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C:Accession: T46188
R:Benes, V.; Rechmann, S.; Borkova, D.; Ansoorge, W.; Mewes, H.W.; Lemcke, K.; Mayer, K.H.
submitted to the Protein Sequence Database, January 2000
A:Reference number: 223014
A:Accession: T46188
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-773 <BEND>
A:Cross-references: EMBL:AL133248
A:Experimental source: cultivar Columbia; BAC clone T8H10
C:Genetics:
A:Map position: 3
A:Introns: 64/2; 146/1; 176/2; 192/3; 223/2; 259/2; 300/2; 484/3; 507/2; 552/2; 625/1; 6
A>Note: T8H10.120

Query Match 34.5%; Score 1458.5; DB 2; Length 773;
Best Local Similarity 36.6%; Pred. No. 1.9e-107;
Matches 294; Conservative 157; Mismatches 249; Indels 103; Gaps 18;

Qy 21 MSSPRAIDGSDFTVNGHSLDVPENIVASPSPIYSIDKSPVSGCFVDPADSEDPDRHV 80
Db 3 ITSNSVQNDNJVVOGKTLITKIPDNIILTP-----VTGNGFVSGSFIGATEQSKSLHV 57

Qy 81 VIGIKLKDREMSIFREFKWTTHWGRNGDPLESETOLVILEK-----SDSGRPVY 132
Db 58 PFIGVLEGRFRCCKRFKIMMTQKRGSCGKDIPLETQMLKSKDEYVNGDADATVTT 117

Qy 133 FLPLPVEGPFRTSIQGDDEYVGVCSGSSKVVDAFSFMSMLYLHAGDDPFLVKEAMKI 192
Db 118 VFLPLEGGFRAVLQNGENKEIEICFESGDKAVETSGQTHLVYVHAAGTNPFEVIRQSVKA 177

```

```

Qy 193 YRHTGTFRLLEKTEPRGIVDKRGKCTWDAFYLYVHPGVIIEGVRHLNDGCGPRGLYLD 252
Db 178 VERHMQTFHNRKRLKPLSPFLDMFGCTWDAFYLYVHAEGBELKSLSEGGPRKRLYLD 237
Qy 253 DGMOSIGH---DSDPITKGMNQTVAEOMPCRLKFOENYKFRDVPNPKATPRAGQK- 308
Db 238 DGMQOLENKEKDCNVCQ-----GAQFATRLVGIKEKNAKQ-----KKDQK 280
Qy 309 ----GKAFIDELKGEFKVEHYVYVHALLCGYVGLRPOVGLR--EARYIOPVLSPLG 362
Db 281 TOYSGLSKVYDNAK -ORHNVKQYAAHLAGYWGKVPRAASGMEHDSALATAYVQSPVL 339
Qy 363 MPMEDLAVDKIVLHKGLVLPPEAEEMYGSLAHLEKVIDGYKIDVILHLEMLCEDYGG 422
Db 340 GNOPIVMDLSLAVHGLVNPVKVFNFNELHSHYLASCGIDGVKVVONITETLGLGGLG 399
Qy 423 RVDLAKAYKAMTKSIKHKFGKGVASMEHCNDMFELGTEATSLGRVGDDEFCPTDPSGD 482
Db 400 RYSLTRSYQDALEASTARNFKNKNGCISCHCHNTDGLY-SAKOTAIYRASDDIYPRDPAS- 457
Qy 483 PNGTFWLGCHVHYCAN---DSLAMGNFTHPDMFQSTHPCAFHAASRAISGPIYVS 539
Db 458 ---HTHIASVAVNLTFLGEFQPDMDHSLHPTEAYHAARAAGCAIYVS 507
Qy 540 DSYGKHNPDLKLVLPDGSILRSEYVALPTRCLEFEDPLHNGETMLKTNLKKFTGVIG 599
Db 508 DKPGNINFDLRLKLVLPDGSILRSEYVALPTRCLEFEDPLHNGETMLKTNLKKFTGVIG 567
Qy 600 AFNCGCGGCRETRRQCSQVSKRYTSKTNPKDI-----EMHSGENPISIGVKTF 650
Db 568 VFMCGAGMKETKKNQIHDTSPGTLGSLRADADLISVAGEDM-SGDS----- 617
Qy 651 FALYIQAKKLLSKPSQDLIALDPFEFELTVSPVTKLIQTSLSHAPIGLVNMLNTSG 710
Db 618 -IYVAYRSEGVRLPKGASIPITLKVLEYELFHISPL-KEITASISFAPIGLDMNSSG 675
Qy 711 AIQSVYD-----YDDLSS-----VEIGVKGCGEMVRASKPRAC 745
Db 676 AIESIDINHTDKNPFPEFGEISSASPALSDNRSPTALVSVSGGCRGAYSSORPLK 735
Qy 746 RIDGEDVGFKYDQD-OMVVQVVP 767
Db 736 AVESTETDFTYDAEVLVTLNLP 758

RESULT 5
C96599
protein F20N2.14 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: C96599
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Malt, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tello
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A66141; MUID:21016719
A:Accession: C96599
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1170 <STO>
A:Cross-references: GB:AE005173; NID:g8778496; PIDN:AAF9504.1; GSPDB:GND0141
A:Gene: F20N2.14
A:Map position: 1

```

```

Query Match      27.0% Score 1140; DB 2; Length 1170;
Best Local Similarity 33.3%; Pred. No. 7.2e-82;
Matches 259; Conservative 135; Mismatches 208; Indels 176; Gaps 20;

QY      26 AIDSDFTVNGHSPSLDVPENIVASP-SPTYSIDKSPVSGCFVGFDASEPDSSRHVNSIG 84
      8  SVTSDSLVLVGLGHVRLHGVPEENVLVTPAGSNLID-----GAFIVTSQDTSNHFVSIF 61
      85 KIKIRIWMISIFREKVMWTTTHVGNRRNGDLESETOVLILEKS-----DSGRPVFLIP 136
      62 KLELRFRMCFVREFKLMMWTQRMGTNGKEIPCTEOTPLVLEANGDSPLGGRDSSSTVFLPL 121
QY      137 IVEGPFRTSIQPGDDDFEYCVESGSSKYVDASFRRSMUYLHAGDDPFLVKEMKIV--- 193
      122 ILBEDFRAVLQGNANELEICLESDDPLVDQPEGSHLVFAAGSDPFDVITKAVYVML 181
      194 RTHLGT-----FLLEKTPPGIYDKRGKCTWDAFYLVHPQSYIEGVRLVD----- 241
      182 KSQKLTSLCPNFFIR-----PDMINMGCTWDAFYNNVYAKDKQGLSNCDLTKPAL 235
QY      242 -----GGCPRGVLVIDDGMSIGSDHDSPITKEANOVTAGEQMPRCILKROENKFRD 294
      236 ILCSLKAGGYPRKYIYIDDGMSVMDETSVEFNADN-----AANFANRLTHLKENKRPQ- 290
QY      295 YVNPKATGPRAGQKGMKA-----FIDELKEFKVEHYVNHALCGTYGGGLRPQVP 345
      291 -----KDKGEHNRVDDPSLSLGHVITDIKSN-NSLKYVYVNHALIGYGGVPRGVS 340
QY      346 GLP---EARVLDIPLVSPGLQMTMEDLAVDKYVHKYGVLPREKAEEMVBEGLNHLKLVGD 403
      341 GMEHYESKVATPVSSPGVMSSENCCLDSITKNGCLGLVAPKEVPSFYNDLHSTYASVGD 400
      404 GVKIDVYIHLLEMLCEDYGRVDLAKAYYKAMTKSINKHKGNGVIASMENCDMFMFLGTE 463
      401 GVKVDVQNTLETLTGAGHGRGVYLAKKYHQDALEASISRNPRDNGIISCMSHNIDGLT-SAK 459
QY      464 AISLGRVDDDWCTDPSGSDPRNGTFFLQGCCHVYHCAN--DSLAMGNFNIHPDMDPQSTPR 520
      460 KTAIVRASDDTWPBRPAS-----HTIHLSAVAYNTLFLGFEPQPDMDMFSLHP 508
      521 CAAFHSAASRAISGPIIYVSDVGNHNFDLKKLVLPDGSILRSEYVALPTRDCLFEDPLH 580
      509 MAEYHAARAVAGGCAIYVSDKRGQHDNFLKRLVLRDGSILRAKLPGRPTRELTY----- 563
QY      581 NGETMLKIMNLKFLPGVIGAFNCGGGCMCRFRRNQCFSQYSKRTYSKINPDIEMHSGE 640
      564 -----LPRD----- 567
      641 NPISIEGVTFEALYUQAKKLILSRPSODLIDALPFEFELITVSPYTKLIQTSLHAFPI 700
      568 -----ISLPYTLMPREVEFTVVPVVKFEPSDGS-KEAPV 599
QY      701 GLVNMMLTSGAIOGVYDDDLSS--VEIGKCCGGEKRVYAS-KKPACRKLIDGEDVQFK 755
      600 GIMEMFVNGGAIVLSRYDDEGTGFYVRMKILRSGLVGYVSSVARRPRSVTYVSDSDEYR 657
      657 -----LPRD----- 657

RESULT      6
T09530
probable seed inhibition protein - chickpea (fragment)
C|Species: Cicer arietinum (chickpea, garbanzo)
C|Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 20-Jun-2000
C|Accession: T09530
R|Cervantes, E.
submitted to the EMBL Data Library, February 1996
A|Reference number: Z16718
A|Accession: T09530
A|Status: preliminary; translated from GB/EMBL/DBJ
A|Molecule type: mRNA
A|Residues: 1-357 <CERN>
A|Cross-references: EMBL:X95875; NID:q1212811
A|Experimental source: germinating seed
A|Genetics:

```

A:Gene: slp

```
Query Match      11.6%: Score 490.5; DB 2; Length 357;
Best Local Similarity 32.5%: Pred. No. 4.4e-31;
Matches 127; Conservative 55; Mismatches 154; Indels 55; Gaps 11;

QY 187 KEAMKIVTNLGTERLLEBEKTRPGIVDKFEGMCTWDAFYLTGNPGVIEGVRLHLDGSGCP 246
      ::::: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3 QQAVKAVEKHNGOTFHNHKKRKRVPSFLDFMGCTMDAFYTDVTAEGVEGKLKSLSEGTPR 62

QY 247 GLVLDGCMOSIGHDSDFITKEGMQTYAGSGMPCRLK-----FOENVK--FRDYVMPK 299
      ::::: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 63 RFLIA-----TWGNKLKKVQSKSGSGCVVQEGNQLGLGWIRDA 101

QY 300 ATGPAGGOK-----GMKAFIDELKEGFEFTVEHVYMAHLCGYGGLRPPVRLPE--AR 351
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 102 NSKKNNGQNDGRIPLRKILVDGK -KHNNKDYVYMAHLAGYWGCVRAATGMEHYDTAL 160

QY 352 VIQPLVSPGLQTMEDLAVDKIVLKVKGLVPRPEKAEEMEGSHALEKVGIDGVYIDYIH 411
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 161 ATSAVTRS--TWNPDPDIWMSDLAVGLRLYHFRGFTSNE--THAALASGVGVAVVDQOT 217

QY 412 LLEMCEGYGGRVDAKAYUKKMTSKINKHFEKGNVIA SMEHCNDFMELGTETALISRGVY 471
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 218 LIRPLVLDTVYESRLHAHAIIMRLRLPLVTLINNCGIACMCNTEGIV-SAKQTAIVRAS 276

QY 472 DDFWCTDPSGDPNGTFTWLOGCHMVH--CANDSLMGNFIHPQDMFOSTHPCAAFHAA 528
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 277 DDFYPHDPA-----HTHISSAYANSFLGEPFQPMQDMFSLHPAAEYHAAA 325

QY 529 RAISGPIYVDSVSGKHNFDLKKIIVDPGS 559
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 326 RAIGCGQFMSVISOATTLILFLKSLVLDGS 356

RESULT 7
T01717
hypothetical protein A_I6002N01.5 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 24-Mar-1999
C:Accession: T01717
R:Scheel, P.; Magg1, L.
submitted to the EMBL Data Library, June 1997
A:Description: The sequence of A. thaliana I6002N01.
A:Reference number: Z14407
A:Accession: T01717
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-711 <SCH>
A:Cross-references: EMBL:AF007269; NID:g2191126; PID:g2191156
A:Experimental source: cultivar Columbia Cigenetics:
A:Map position: 4
A:Introns: 5/2; 25/3; 103/3; 126/1; 174/2; 240/1; 284/1; 295/1; 333/2; 349/2
A>Note: _A_I6002N01.5

Query Match      11.2%: Score 471.5; DB 2; Length 371;
Best Local Similarity 26.9%: Pred. No. 1.5e-29;
Matches 131; Conservative 71; Mismatches 112; Indels 173; Gaps 18;

QY 106 VGRNGDLESEFQIVILEKS-----DSGRPYFLPLIVSGSPRTSIQRPDDDFVNYC 157
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 27 MCTNGKEIPCETQFLIVEANKSGSLGGDESSVYVFLILGDFRAVYQGEANLELIC 86

QY 156 VSSGSKVYDAFRSLYLHAGDDPALVKEMKIVKRTLGTFRLLLEKTPPGIVKFGM 217
      ::::: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 87 LBSGR-----LTOLARFEL--KLFMDMLNWFQM 113

QY 218 CTWDAFYLTIV-----HPQGVIEGVRLHLDGSGCPGLVLDGCMOSIGHDSDPITKEGM 270
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 114 CTWDAFYKRVYLDCCDLTKRAIILCSLK--AGVYPRKFVYIDGMSVGMDE---TSVEF 167
```



```

OY      623  KRYVTS---KTNP-----KDIEMSGENPISIEGVKFTALVQAKKILSKPS 667
          |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
Db      490  KPTAYAVMLDYNVRIILVTASLDGLIKFWDFENKGNLIDSLDGSSTITHAIYQHSPLVAVAC 549
          |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
OY      668  QDDLDI-----ALDPFEFE-----LITVS-----PVTKLIQ- 692
          |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
Db      550  DDEIRIVDVQTRKIVRELIMGHSNRILTSFDESDGRMLVTASLDGIRTMWLDLPGHLIDS 609
          |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
OY      693  -----TSLHFAPIGLVNMLNTSGAIQSDVYDDDLSSVEIGVGCCEGEMRFASKRRPAC 745
          |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
Db      610  ISTRVCTSLTFAPETG-----DYLAFTTHDVQGISLMTNLSMFKHVSSTKAL 655
          |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
OY      746  RIDGEDYGFRTYDQDQMYVYQWPPIDSSSGISYIE 781
          |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
Db      656  RLDD-----VVEVSAPSVSGEKGISYVE 678
          |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:

RESULT 10
T42245
probable polypeptide N-acetylglactosaminyltransferase (EC 2.4.1.41) - Caenorhabditis el
C:Species: Caenorhabditis elegans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jul-2000
C:Accession: T42245
R:Hagen, F.K.; Nehrke, K.
J. Biol. Chem. 273, 8268-8277, 1998
A:Title: cDNA cloning and expression of a family of UDP-N-acetyl-Dgalactosamine:polypep
A:Reference number: 722126; MUID:98192620
A:Accession: T42245
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-623 <HMG>
A:Cross-references: EMBL:AF031835; NID:g3047190; PIDN:AAC13671.1; PID:g3047191
C:Gene: g1y-5
C:Keywords: glycosyltransferase; N-acetylglactosaminyltransferase
C:Superfamily: polypeptide N-acetylglactosaminyltransferase
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match      2.7%; Score 114; DB 2; Length 623;
Best Local Similarity 21.7%; Pred. No. 0.84;
Matches 80; Conservative 55; Mismatches 122; Indels 112; Gaps 21;

OY      220  WDAFLTYVHPGVIIEGV-RHLVDGCGPGLYLIDGQSIGHSDSDPIRKEGNQTVAGEQ 278
          |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
Db      188  MSVILRTVH--SVLETPDHLE-----EVLAVD-FSDMDHTKRL-ETWISQ-FGGRV 237
          |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
OY      279  MPCRLIKFOENYKFFDYVNPKRATGP-----RAGQKGMAFIDELKGEFTV----- 324
          |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
Db      238  KILRKEKKEGLIRALRGAATAATGEVLTYIDSHCCMGSMWEPILDRIKRDPITYVVCVI 297
          |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
OY      335  ---EHYVVMH-----ALCGYWGGLRPQVGLPE-----ARVIVPLSPGLQMTMED 367
          |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
Db      298  DIVDNTFEYHNSKAYFTSVSGFDWGLQFMNHSIPEDRKNTKTRPIDIVRSP---TAAQ 353
          |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
OY      368  --LAVDKIVLKHVG-----LVPPEKAEMEYEGELHNL 397
          |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
Db      354  GLFSIDKEYFEELGYDPGFDIWMGENIELSFKIMCGTLEIVPCSHVGHVFRKRSPIK 413
          |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
OY      398  EKVGIDGVYDIVIHLEMLCEDYDGRVLDLAKAYYKAMTKSINKHFKGNGVIAISMEH---- 453
          |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
Db      414  WRTGVNVLKRNRIRLAEVWLDY-----KTYI---YERINNOJGDGDIDSSRKKLRED 463
          |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
OY      454  --CNDP-----MFLGTEAISLGRV---GDFEWCSD--PSGDPNGFTWLOGCHMV 495
          |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
Db      464  LGCKSFKNWYLDNIYPELFPVGSVAKGEMRNAGKRNQCIDYKPSG--GKTVGMAYQCH-- 519
          |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
OY      496  HCANDSLMVA 504
          |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
Db      520  NOGGNOYWM 528
          |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:

```

hypothetical protein YDR419w - Yeast (*Saccharomyces cerevisiae*)
C:Species: *Saccharomyces cerevisiae*
C:Date: 22-Aug-1996 #sequence_revision 06-Sep-1996 #text_change 23-Mar-2001
C:Accession: S69702
R:Dietrich, F.S.
submitted to the EMBL Data Library, August 1995
A:Description: The sequence of *S. cerevisiae* lambda 3641 and cosmids 9461, 9831, and
A:Reference number: S69555
A:Accession: S69702
A:Molecule type: DNA
A:Residues: 1-632 <DIE>
A:Cross-references: EMBL:U03007; NID:927685; PIDN:AAB64856.1; PID:927690; MIPS:YDR4
C:Genetics:
A:Gene: SCD:RAD30
A:Cross-references: SGD:S0002827; MIPS:YDR419w
A:Map position: 4R

Query Match	2.7%	Score 113.5	DB 2	Length 632
Best Local Similarity	21.2%	Pred. 0.94		
Matches 113	Conservative 55	Mismatches 197	Indels 169	Gaps 23

QY	277	EQMPCRLK-----	FOENYKFRDYVNEKATGPAGQGMKAFIDELKGEFTY	324
Db	39	EQMNGSLKEDPPVVCQVMNSIIAVSAAKRGISRMDTIOEALKKCSNLIPIHTAFVKKG	98	
QY	325	EHYVYVHALGCGYMGGLRPQVPGLEPEARKYIQVLSPGIQMTMEDLVAKYLVHVGVP--	382	
Db	99	EDFMQYHGGCGSW-----	VODPAQIASED-----	HKVSEPYR 132
QY	383	--PKEAEENYEGHLAHLEKVGIDGVKIDVIHL--	LEMLCEDYGGRVDLAKAYYKAMTKS	437
Db	133	RESRKALKITFSACDIVERASIDEVFLDGRICFMNMLPNEVELTGDK--	LKDALSN 189	
QY	438	INKHFKNGVYASMEHCNDEMFLGTETAI-SLGRVDDPWCITDSESGDNFTW-----	488	
Db	190	IREAFIGSNY-----	DINSHLPLIPEKIKSLKEGDFV--	NPEGRLDITWDDVYIALG 241
QY	489	-----	LOGCHWVCANDSLMNGNFIHPDW-----	DMFOSTHPCAFH 525
Db	242	SOVCKGIRDSTIKDILGYTTSGLSSTKNVCKLASNKKPDAQTIVKNDCLLDLPDCKFE	301	
QY	526	AASRAISGSPITVSDSVGKHNFDLKLVLDP-----	DGSILR 562	
Db	302	ITSEFTLGGVL-----	GKELIDVID--	LPHENSIKIHIRETPMDNAGOLKEFLDAVKQ 352
QY	563	SEYIALPRLDCLFEDPLNGETMLKINLTK-----	FTYIGAFNCGGG--	607
Db	353	SDY-----	DRSTSNIDPLKTADLAELFKLSGRGYGLPSSRPVYKSMNSKNLNGKACNS	408
QY	608	-----	MCRE-TRRNOCFSO-YSKRVYSKT-----	NPKDIEHWSGENPISIEGVKT 650
Db	409	IYDCISLMEVFCALETSLTIQDLEQYKNIIVIPRVYSISLTKSYEYRKSPVAKGIN-	467	
QY	651	FALVLYQAKKIL--SKPSQDIDALDFEFELITVSEVYTKLIOTSILHFAPIGL	702	
Db	468	-----	FGSHELLKVGKIKFVTDLDI-----	KGKNSYVPLTKLSMTITNFIIDL 511

RESULT	12
164118	1.4-alpha-glucan branching enzyme (EC 2.4.1.18) - Haemophilus influenzae (strain Rd K
C:Species:	Haemophilus influenzae
C:Date:	10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession:	164118
R:Flatschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage	
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.T.; Glodek, A.; Kelley, J.M.; Weidman	
; D.M.; Brandon, D.C.; Fine, L.R.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.	
Science 269, 496-512, 1995	
A:Authors:	Gnehm, C.U.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter
A:Title:	Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number:	A64000; MUID:95350630
A:Accession:	164118

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-730 <TIG>

A:Cross-references: GB:U32815; GB:LA2023; NID:g1574818; PIDN:ACC23004.1; PID:g1574820;

C:Superfamily: 1,4-alpha-glucan branching enzyme

C:Keywords: glycogen/starch biosynthesis; glycosyltransferase; hexosyltransferase

Query Match 2.5%; Score 105; DB 1; Length 730;

Best Local Similarity 17.5%; Pred. No. 5.6; Mismatches 203; Indels 272; Gaps 29;

Matches 117; Conservative 75; Mismatches 203; Indels 272; Gaps 29;

68 VGFDEPSRHHVSGIKLDIRFMSIFRKYMTTHWGRNGC---DLESETOYILEK 124
 138 VNRRLAPNARRSIYGF-----NYMGRKRIPRPHSSGWELFLPK 181
 125 SDGRPYFLPLIVEG-----PFRSTIQPGDDDFVDCVSGSSKVVDAFSRMLYLH 177
 182 ASIGQLYKELLIDCHNLRLKADPFASFSSQLRDTASQVSALPNVEMTEARKKA---N 237
 178 AGDPRALYKEMAKIVRTILGTRRLLEKTRPVIVKFGCTDAFYLYHPGVIEGVR 237
 238 QGNP-----ISITYEHLGWSR-----RNLENNF-WLDYDQIADDELIPYKEMGFT 282
 238 HL-----VDGCRPGLVLDGCMOSIGHSDPTKSGNMQTVAGSQMPRLKLFQE 288
 283 HIEFLPLSEPRFGSW-----GYRPLGLYS-PTSRRGSPRAR-----RLVK--- 323
 289 NKFRDYVPKATGPRAQCKMKAFIDELKEERTEVHYVHALCGYWGRLRPQVGLP 348
 324 -----RAHEAGINVLIDWVGHPERSDTH----- 346
 349 EARVIOPLVSPGLQMTMEDLAVDKIVLHKVGLVPRPEKAEEMTEGLAH----- 366
 347 -----GL-----VAEDGALY-----EHEDPREGQHDWMTLLIYNTGRNE 381
 397 -----LEKYGIDGVKIDVILHL-----EMLCEDGGRVLDL-AKAYYKA 433
 382 VKNFSSNALYMERGVGIRYDVAASMIYRYSRAEGWIMINQYGGKRLNLAIEFLKH 441
 434 MKYSIKKHKNGVIAEMHCNDFMFLGTEAISLGRVGDFTWCTDPSGDNPTGFWLQGGH 493
 442 TWKKI--HSEMGALISIAESTSFAGV--THPSENGGLGFEF-----KMNNG-- 484
 494 MYNCADSLMNGFIHPDMDMOSTHPCAFHNASRAISGRITYVDSVSKHNFDLKL 553
 485 -----WMNDTL-----AYMKLDPIYROYHNKMTGFGVYQY 515
 554 VLPDGSILRSEYVALPTRDCLFEDPLN-----GETMLKIMLNKFTGVIGAF--- 601
 516 -----SENFVRLS--HDEYVHGKYSLLGKMPGDTWQKFAHLRLAYGYIMNGYRPG 563
 602 -----NCQ-----GGGWCSETRR-----NOCFSQYSKRVTSKT 629
 564 KLLFMGNEFAQREWMYBESLDWFLLDENIGGKMGKVLKVLNDLQIKNRPLELDN 623
 630 NPDIEM 636
 624 SPEGFDM 630

RESULT 13

T19997

hypothetical protein CA7D12.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T19997

R:Gatedsky, S.

A:Reference number: T19209

A:Accession: T19997

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-3944 <ML>

A:Cross-references: EMBL:669902; PIDN:CAA93765.1; GSPDB:GN00020; CESP:CA7D12.1

A:Experimental source: clone CA7D12

C:Genetics:

A:Gene: CESP:CA7D12.1

A:Map position: 2

A:Introns: 46/3; 308/3; 408/2; 449/3; 820/3; 946/3; 1069/1; 1634/3; 1737/2; 2009/2; 2

Query Match 2.4%; Score 103; DB 2; Length 3944;

Best Local Similarity 18.5%; Pred. No. 1.1e+02;

Matches 136; Conservative 90; Mismatches 230; Indels 280; Gaps 34;

80 VVSGIKLDIRFMSIFRKYMTTHWGRNGDLESETOYILEKSSGRPYFLPLIVE 139
 2931 VVSYGHVKIQLANLV-----EIEESTDIRT----- 2957
 140 GPRFTSIQPGDDDFVDCVSGSSKVVDAFSRMLYLHAGDPRALYKEMAKIVRTILGT 199
 2958 -----ALLEAPSNKVDQALM-----GD-----MKSLMKYFR----- 2983
 200 FLLEKTRPGIYDKFGCTWDAFYLYTHPQVIEGVRLHYDGCGRPLVLD-DGQST 238
 2984 -----NRTPTSDDMGFEVSTWYDMRQIH-----GMMLQRFEEYWDKY 3020
 259 GHSDSDPTKSGNMQTV-----AGQMPORLLKFOENYKFRDYVPKATGPRAQCKMKAFI 314
 3021 GLN---VAATG--NOSTIVPHISMAQDLAVAKHAKNNGFHLT-----KDL 3062
 315 DELKEGFEVHYVNHALCGYWGRLRPQVGLPEARVIOPLVSPGLQMTMEDLAVDKIV 374
 3063 NKLAG--LTAIRPMADQDKVCTYKTLRDMANSADEVKNELCEALEV--LEQVRIDDLQ 3120
 375 -----LHKVGLVPRPEKAEEMTEG-----LHAHLKVGIDGVKIDVILHLE- 414
 3121 KDQVAAALYHRANIHISV--LDOAENADYTFSAASQVLDLQNSVTTGTGIKLAKNMGHNLKYR 3179
 415 -----MLCEDG---GRVLDKAKAYYKA-----MTKSIN--KHFGNGV 447
 3180 FSTTYCKEKGNNFGQALACVFIARVNDIDAKRPIAKILMLSKHLNCGSHENVNRY 3239
 448 IASMEH-CNDFMFLGTEAISLGRVGDFTW--CTDPSGDNPTGFWLQGGHMYNCADSL 502
 3240 IKKQLSLNLFNL-----YMLPOLVTDVRYKPNENFVLICKMA----- 3279
 503 WMGNFTHPDMFQSTHPCAFHNASRAISGRITYVDSVSKHNFDLKLVLDPGSILR 562
 3280 -----AAHPLOVFYHIREAVSVDDI-----DSVLE 3304
 563 SEX-----YALPTRCLEFEDPLHNGETMLK-----WNLNKFYGVIGAFNCGGGW-- 608
 3305 EDTTDEOMSDVDEDFADDPF--DRILKICLKRPDIRVHRVLYKLDEENETWVE 3362
 609 -----CRETRNOCFS-----OYSKRVSTKTNPKDIEHNGENPISIEGVKTP 651
 3363 RHLRHAICLKQDFKDFSEDMATFNMDSYEDVTMT-----LWR-----KQLEED 3410
 652 ALYLYQ-----AKLLISKPSQDIDALDPPEFLITYSPYTKLIQTSILHAP 700
 3411 LVYFOQNYNLDFLEIRNKRKMIVTKGCMGVEKQSIIMEKELSVFTERPACMODEPFDV-T 3469
 701 GLVNLMTSGAIOVD 716
 3470 NMTNMVVSQDLIHAVD 3485

RESULT 14

HM1VF7

hemagglutinin precursor - Influenza A virus (strain A/ruddy turnstone/NJ/47/85 [H4N6])

N:Contains: Hemagglutinin HA1; Hemagglutinin HA2

C:Species: Influenza A virus

C:Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 18-Sep-1998

C:Accession: G34214

R:Donis, R.O.; Bean, W.J.; Kawooka, Y.; Webster, R.G.

Virology 169, 408-417, 1989

A:Title: Distinct lineages of influenza virus H4 hemagglutinin genes in different region

A:Reference number: A34214; MUID:89204912

A:Accession: G34214

A:Molecule type: genomic RNA

A:Residues: 1-564 <DON>

C:Genetics:

A:Map position: segment 4

C:Superfamily: influenza virus hemagglutinin

C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond

F:1-16/Domain: signal sequence #status predicted <SIG>

F:17-343/Product: hemagglutinin HAI #status predicted <HAI>

F:344-564/Product: hemagglutinin HA2 #status predicted <HA2>

F:534-550/Domain: transmembrane #status predicted <TM>

F:18,34,178,310,497/Binding site: carboxylate (Asp) (covalent) #status predicted

F:26-480,64-79,151-487,295-319/Disulfide bonds: #status predicted

F:553,560,563/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 2.4%; Score 102.5; DB 1; Length 564;

Best Local Similarity 18.6%; Pred. No. 5.9;

Matches 110; Conservative 61; Mismatches 188; Indels 233; Gaps 26;

QY 7 NGGSNNVSFDGLN-----MSSPPAIDGSDFTVNGHSFLSDVPEENIVASPSPTISDKS 60

DB 80 NGALGSPGCDHLNEMDVFIERPDAVD-----TCYPRDVPD-----YQSIRSI 123

QY 61 PVSVCFCFPGFASPSDSRHVYSICK--LKDIRFMSIFRFYVWMTTHVGR----- 108

DB 124 LANNGKE-EFLTAERQMTVAVQNGSKGACRRANVNDFFR-RLNMLTKSDGANVPLQNLTK 181

QY 109 -NGGDL-----ESETOIVLEKSDSGR-----PYVFLPIVEGP-- 141

DB 182 VNNGDYARLYIMGVNHPSTDEQFNLKYNKMPGRVYVSTKTSQTSVVPVIGSRPLRGQSG 241

QY 142 ----FRSISQPGD-----DDEVDVCYESGSS 163

DB 242 RISFWTIVEPDLIVFTIGNLILAPRGHYKLNSQKSTIINTAVPIGSCVSKCHTNRGS 301

QY 164 KVDASFSRMLYLAHGDPPALVKEAMKIVRTHLGTFPLEKTPPGIIVDFGQMTWAF 223

DB 302 ITTTRKPFONISRIISIGDCPKYVKGSLKLA*---GMRNIPKATRGLP----- 346

QY 224 YLTVAPOGVIEGVRLHVLGGCPGLVLIDDMQOSIGHSDPITKEGMNQTVAGEQMPCL 283

DB 347 -----GALNG-----FIENMGQGL----- 360

QY 284 LKFDENKFRDYNKATGPRAGOKMAFTIDELKGEFTVEHVVVWHAALCGYWGGLRPQ 343

DB 361 --IDGWYGF-R-HQNAEGTGTADLKSTQALDQINGKLRL----- 398

QY 344 VPGLEAVIOPVLSFGIOMMEDLAVDKIVLHKYGLVPRPEKAEEMEGSLAHLEKVGID 403

DB 399 -----LEKTNEK-----YHQI-----EKEFQVBERIODLEKY-VE 428

QY 404 GVKIDV-----IHLLEMLCEDYGGRYDLAKAYYKATKSI-----NKHFGNGVYASMEH 453

DB 429 DTKIDIMSYNNELVALLENQH--TIDVTDSSENNKLFERYVRQLRENAEDKGGCFEIRHQ 486

QY 454 CNDENFLCTEASISGRVDDFWCTDPSGDPNGTFWLGCGHAYHCAND-SLWM 504

DB 487 CDNNICI---ESIRNGTYDHDIVYDEAI---NNRFQIQGVKILQGYKDIILWTI 532

RESULT 15

S74819 extracellular solute-binding protein - *Synechocystis* sp. (strain PCC 6803)

N:Alternate names: hypothetical protein slr1740

C:Species: *Synechocystis* sp.

A:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 28-Jul-2000

C:Accession: S74819

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, O.; K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas

DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocys*

S.

A:Reference number: S74322; MUID:97061201

A:Accession: S74819

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-582 <KAN>

A:Cross-references: EMBL:D90909; GB:AB001339; NID:91652844; PIDN:BA117780.1; PID:9165

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Genetics:

A:Start codon: GTG

C:Superfamily: oligopeptide transport system permease appa

Query Match 2.4%; Score 102.5; DB 2; Length 582;

Best Local Similarity 25.5%; Pred. No. 6.2;

Matches 55; Conservative 20; Mismatches 68; Indels 73; Gaps 11;

QY 3 PSFNGGSNNVSFDGLN-----DMSSPPAIDGSDFTVNGHSFLSDVPEENIVASP 51

DB 97 PSVENGG---VAADGLSVTWKIKPDVLMWDGQPFSAEDVAF*---YKFLSD-PKTGATST 149

QY 52 SPYTSI-----DKSPVSGCFVGFDASEPDSRHVYSICKLKDIRFMSIFRFKVMWTTHW 105

DB 150 GYEAIAKVEALDKNTVYK1-----TFKEPNDA-----WFLPF 181

QY 106 VGRNGDLESETOIVLEKSDSGRPYVFLPIVEGPR-TSIOPGD----- 150

DB 182 VGSSEMLLPQHTYKDFVEKARQAP-ANLLPIGIGPRVYTSFKFGDVLVYVNHRYDRK 240

QY 151 -DDEVDVCYESGSSKVDASFSRMLYLAHGDPPAL 185

DB 241 NIGFOQVEIKGGG---DATSAARAVLQTDGADPAL 272

Search completed: November 30, 2001, 09:41:50

Job time: 89 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 30, 2001, 09:40:21 ; Search time 70.16 Seconds

(without alignments)
31.673 Million cell updates/sec

Title: US-08-846-234-1

Perfect score: 181
Sequence: 1 FGCWTDARFLYTHPGCVIRGVRHLYDGGC 30

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq.1101.*
1: /SID8/gcgdata/geneseq/geneseq/AA1980.DAT:*
2: /SID8/gcgdata/geneseq/geneseq/AA1981.DAT:*
3: /SID8/gcgdata/geneseq/geneseq/AA1982.DAT:*
4: /SID8/gcgdata/geneseq/geneseq/AA1983.DAT:*
5: /SID8/gcgdata/geneseq/geneseq/AA1984.DAT:*
6: /SID8/gcgdata/geneseq/geneseq/AA1985.DAT:*
7: /SID8/gcgdata/geneseq/geneseq/AA1986.DAT:*
8: /SID8/gcgdata/geneseq/geneseq/AA1987.DAT:*
9: /SID8/gcgdata/geneseq/geneseq/AA1988.DAT:*
10: /SID8/gcgdata/geneseq/geneseq/AA1989.DAT:*
11: /SID8/gcgdata/geneseq/geneseq/AA1990.DAT:*
12: /SID8/gcgdata/geneseq/geneseq/AA1991.DAT:*
13: /SID8/gcgdata/geneseq/geneseq/AA1992.DAT:*
14: /SID8/gcgdata/geneseq/geneseq/AA1993.DAT:*
15: /SID8/gcgdata/geneseq/geneseq/AA1994.DAT:*
16: /SID8/gcgdata/geneseq/geneseq/AA1995.DAT:*
17: /SID8/gcgdata/geneseq/geneseq/AA1996.DAT:*
18: /SID8/gcgdata/geneseq/geneseq/AA1997.DAT:*
19: /SID8/gcgdata/geneseq/geneseq/AA1998.DAT:*
20: /SID8/gcgdata/geneseq/geneseq/AA1999.DAT:*
21: /SID8/gcgdata/geneseq/geneseq/AA2000.DAT:*
22: /SID8/gcgdata/geneseq/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	181	100.0	30	19	AAW53567
2	181	100.0	30	20	AAV17422
3	181	100.0	784	19	AAW53570
4	181	100.0	784	20	AAV17417
5	158	87.3	758	21	AAV70978
6	148	81.8	799	19	AAW57886
7	148	81.8	799	20	AAV30142
8	146	80.7	587	19	AAW57888
9	146	80.7	780	22	AAW30144
10	146	80.7	781	19	AAW57887
11	146	80.7	781	19	AAW57887

12	146	80.7	781	20	AAV30143	Amino acid sequenc
13	146	80.7	781	22	AAW49400	Soybean raffinose
14	144	79.6	572	20	AAV32075	Rapeseed raffinose
15	144	79.6	777	20	AAV32074	Mustard raffinose
16	122	67.4	783	20	AAV32073	Sugarbeet raffinose
17	108	59.7	751	21	AAV70980	Wheat raffinose sy
18	107	59.1	763	21	AAV70977	Rice raffinose syn
19	106	58.6	750	20	AAV17418	Soybean raffinose
20	102	56.4	770	20	AAV70976	Rice raffinose syn
21	98	54.1	841	21	AAV70981	Wheat raffinose sy
22	97	53.6	756	21	AAV70975	Corn raffinose syn
23	71	39.2	10	20	AAV17419	Soybean raffinose
24	68	37.6	150	11	AAV30961	Basic fibroblast g
25	62	34.3	40	14	AAV43287	RCF antagonist [AI
26	60	33.1	40	14	AAV43286	RCF antagonist, bfg
27	60	33.1	43	13	AAV4342	RCF residues 27-69
28	60	33.1	44	21	AAV18542	Immunogenic peptid
29	60	33.1	45	8	AAV71559	Fibroblast growth
30	60	33.1	45	14	AAV43278	RCF antagonist bfg
31	60	33.1	45	21	AAV18551	Immunogenic peptid
32	60	33.1	86	9	AAV81933	Human basic fibrob
33	60	33.1	101	10	AAV90557	rhbFGF mutlein CS10
34	60	33.1	105	10	AAV90558	rhbFGF mutlein CS10
35	60	33.1	114	10	AAV90559	rhbFGF mutlein C14
36	60	33.1	114	22	AAV60696	Human basic fibrob
37	60	33.1	118	10	AAV90560	rhbFGF mutlein C18
38	60	33.1	123	10	AAV90561	rhbFGF mutlein C123
39	60	33.1	129	9	AAV81940	Human basic fibrob
40	60	33.1	129	10	AAV90562	rhbFGF mutlein C129
41	60	33.1	129	10	AAV90564	rhbFGF mutlein CS23
42	60	33.1	132	20	AAV17995	Human basic fibrob
43	60	33.1	134	9	AAV81932	Human basic fibrob
44	60	33.1	134	15	AAV65933	Rat fibroblast gro
45	60	33.1	135	15	AAV5931	Fibroblast growth

ALIGNMENTS

RESULT 1	
AAW53567	
ID	AAW53567 standard; peptide: 30 AA.
XX	
AC	AAW53567;
XX	
DT	06-JUL-1998 (first entry)
XX	
DE	Cucumber raffinose synthase residues 215 to 244.
XX	
KW	Cucumber; raffinose synthase; sucrose; galactinol.
XX	
OS	Cucumis sativus.
XX	
PN	JPI0084973-A.
XX	
PD	07-APR-1998.
XX	
PF	28-APR-1997; 97JP-0111124.
XX	
PR	26-JUL-1996; 96JP-0198079.
XX	
PR	26-APR-1996; 96JP-0107682.
XX	
PA	(AJIN) AJINOMOTO KK.
XX	
DR	WPI; 1998-264858/24.
XX	
PT	Raffinose synthase gene - useful for preparation of raffinose in
XX	transformed plant
PS	Claim 2; Page 16; 26pp; Japanese.
XX	
CC	The present sequence is a cucumber raffinose synthase fragment.
	Raffinose synthase forms raffinose from sucrose and galactinol, has

CC an optimum pH of 6 to 8 and working temperature of 35 to 40 degrees
 CC C, has a molecular weight of 75 to 95 kDa by gel filtration or 90
 CC to 100 kDa by PAGE and SDS-PAGE under reductive conditions and is
 CC inhibited by iodoacetamide, N-ethylmaleimide and myoinositol.

SQ Sequence 30 AA;

Query Match 100.0%; Score 181; DB 19; Length 30;
 Best Local Similarity 100.0%; Pred. No. 1.8e-19;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FGWCTWDAFYLTVPVHPOGVIEGVRHLVDGCG 30
 ||||||||||||||||||
 Db 1 fgwctwdaftytlvhpqvgviegvrhlvdgsc 30

RESULT 2

AAV17422 standard; peptide; 30 AA.

AC AAV17422;

DT 29-JUL-1999 (first entry)

DE Cucumber raffinose synthase peptide SEQ ID NO:1.

XX Raffinose synthase; sucrose; galactinol.

XX Cucumis sativus.

OS JP1123080-A.

XX 11-MAY-1999.

PF 24-OCT-1997; 97JP-0292969.

PR 24-OCT-1997; 97JP-0292969.

XX (AJIN) AJINOMOTO KK.

XX WPI; 1999-340516/29.

PT New raffinose synthase gene - for production of raffinose from
 sucrose and galactinol

PS Example 2; Page 21; 37pp; Japanese.

XX The present invention describes a raffinose synthase, having an activity
 CC of forming raffinose from sucrose and galactinol. The raffinose synthase
 CC gene can be used for expression in a plant for the production of
 CC raffinose. The raffinose synthase can give raffinose from sucrose
 CC and galactinol efficiently. The present sequence represents a raffinose
 CC synthase peptide from cucumber.

XX Sequence 30 AA;

Query Match 100.0%; Score 181; DB 20; Length 30;
 Best Local Similarity 100.0%; Pred. No. 1.8e-19;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FGWCTWDAFYLTVPVHPOGVIEGVRHLVDGCG 30
 ||||||||||||||||||
 Db 1 fgwctwdaftytlvhpqvgviegvrhlvdgsc 30

RESULT 3

AAW53570 standard; Protein; 784 AA.

XX AAW53570;
 AC
 XX

DT 06-JUL-1998 (first entry)

XX Cucumber raffinose synthase.

XX Cucumber; raffinose synthase; sucrose; galactinol.

XX Cucumis sativus.

OS JP10084973-A.

PN 07-APR-1998.

PD 28-APR-1997; 97JP-0111124.

XX 26-JUL-1996; 96JP-0198079.

XX 26-APR-1996; 96JP-0107682.

XX (AJIN) AJINOMOTO KK.

DR WPI; 1998-264858/24.

DR N-PSDB; AAV22250.

XX Raffinose synthase gene - useful for preparation of raffinose in
 transformed plant

PS Claim 3; Pages 17-20; 26pp; Japanese.

XX The present sequence is cucumber raffinose synthase, which
 CC forms raffinose from sucrose and galactinol, has an optimum pH of
 CC 6 to 8 and working temperature of 35 to 40 degrees C, has a
 CC molecular weight of 75 to 95 kDa by gel filtration or 90 to 100 kDa
 CC by PAGE and SDS-PAGE under reductive conditions and is inhibited by
 CC iodoacetamide, N-ethylmaleimide and myoinositol.

XX Sequence 784 AA;

Query Match 100.0%; Score 181; DB 19; Length 784;
 Best Local Similarity 100.0%; Pred. No. 8.2e-18;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FGWCTWDAFYLTVPVHPOGVIEGVRHLVDGCG 30
 ||||||||||||||||||
 Db 215 fgwctwdaftytlvhpqvgviegvrhlvdgsc 244

RESULT 4

AAV17417 standard; Protein; 784 AA.

XX AAV17417;

XX 29-JUL-1999 (first entry)

DE Cucumber raffinose synthase.

XX Cucumber; raffinose synthase; sucrose; galactinol.

XX Cucumis sativus.

OS JP1123080-A.

PN 11-MAY-1999.

PD 24-OCT-1997; 97JP-0292969.

XX 24-OCT-1997; 97JP-0292969.

XX (AJIN) AJINOMOTO KK.

DR WPI; 1999-340516/29.

DR N-PSDB; AAX61238.

in whichs same as on
 us patent 6,146,292

PT New raffinose synthase gene - for production of raffinose from
 PT sucrose and galactinol

XX
 PS Claim 2: Page 25-27; 37pp; Japanese.

CC The present invention describes a raffinose synthase, having an activity
 CC of forming raffinose from sucrose and galactinol. The raffinose synthase
 CC gene can be used for expression in a plant for the production of
 CC raffinose. The raffinose synthase can give raffinose from sucrose
 CC and galactinol efficiently. The present sequence represents raffinose
 CC synthase from cucumber.

XX
 SQ Sequence 784 AA:

Query Match 100.0%; Score 181; DB 20; Length 784;
 Best Local Similarity 100.0%; Pred. No. 8, 2e-18;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FGWCTDAFYLTVHPQVIEGVRLVDGCG 30
 ||||||||||||||||||||||||||||
 DB 215 fgwctdaftytlvhpqvgvrlvdygvc 244

RESULT 5

AAV70978
 ID AAV70978 standard; Protein; 758 AA.

XX
 AC AAV70978;

XX
 DT 09-AUG-2000 (first entry)

XX
 DE Soybean raffinose synthase from clone sfl1.pk125.d4.

XX
 KW Soybean; raffinose synthase; raffinose saccharide;

XX
 KM clone sfl1.pk125.d4; nutritional; soy protein.

XX
 OS Glycine max.

XX
 PN WO200024915-A2.

XX
 PD 04-MAY-2000.

XX
 PF 22-OCT-1999; 99WO-US24923.

XX
 PR 23-OCT-1998; 98US-0105451.

XX
 PA (DUPO) DU PONT DE NEMOURS & CO E I.

XX
 PI Allen SM, Hiltz WD;

XX
 DR WPI: 2000-350754/30.

XX
 DR N-PSDB; AAD00335.

XX
 PT Nucleic acids and encoded proteins involved in the biosynthesis of
 PT raffinose, useful for producing soybean seeds with a reduced raffinose
 PT content and therefore improved nutritional quality -
 XX
 PS Claim 2: Page 47-49; 58pp; English.

XX
 CC The present sequence is a raffinose synthase from
 CC clone sfl1.pk125.d4 isolated from a soybean immature flower cDNA
 CC library sfl1. Raffinose synthase is involved in the biosynthesis
 CC of raffinose and higher homologues in the raffinose saccharide family
 CC from sucrose. The present sequence is useful for reducing the raffinose
 CC saccharide content of soybean seeds which improves the nutritional
 CC quality of the soy protein products derived from them.

XX
 SQ Sequence 758 AA:

Query Match 87.3%; Score 158; DB 21; Length 758;
 Best Local Similarity 86.7%; Pred. No. 1, 7e-14;

Matches 26; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 FGWCTDAFYLTVHPQVIEGVRLVDGCG 30
 |||||||||||||||||||||||||
 DB 202 fgwctdaftytlvhpqvgvrlvdygvc 231

RESULT 6

AAV57886
 ID AAV57886 standard; Protein; 799 AA.

XX
 AC AAV57886;

XX
 DT 23-SEP-1998 (first entry)

XX
 DE Broad bean raffinose synthetase.

XX
 KW Raffinose synthetase; metabolism modification; food additive;

XX
 KM gastrointestinal flora; broad bean.

XX
 OS Vicia faba.

XX
 PN EP849359-A2.

XX
 PD 24-JUN-1998.

XX
 PF 18-DEC-1997; 97EP-0122417.

XX
 PR 18-DEC-1996; 96JP-0338673.

XX
 PA (SUMO) SUMITOMO CHEM CO LTD.

XX
 PI Oeda K, Wantanabe E;

XX
 DR WPI: 1998-324670/29.

XX
 DR N-PSDB; AAV40800.

XX
 PT New nucleic acid molecule encoding plant raffinose synthetase -
 PT capable of producing raffinose, used as food additives with
 PT beneficial effects on gastrointestinal flora

XX
 PS Claim 1: Page 26-29; 44pp; English.

XX
 CC This sequence represents the broad bean raffinose synthetase of the
 CC invention. The raffinose synthetase is capable of producing raffinose by
 CC combining a D-galactosyl group through an alpha (1-6) bond with a
 CC hydroxyl group attached to the carbon atom at position 6 of a D-glucose
 CC residue in a sucrose molecule. The DNA can be used to modify metabolism
 CC of a host organism by introducing into the host organism or its cell so
 CC that the content of the raffinose family oligosaccharides in the host
 CC organism or cell is changed. Raffinose oligosaccharides are useful as
 CC food additives with beneficial effects on the gastrointestinal flora.

XX
 SQ Sequence 799 AA:

Query Match 81.8%; Score 148; DB 19; Length 799;
 Best Local Similarity 80.0%; Pred. No. 5, 2e-13;
 Matches 24; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 FGWCTDAFYLTVHPQVIEGVRLVDGCG 30
 |||||||||||||||||||||
 DB 238 fgwctdaftytlvhpqvgvrlvdygvc 267

RESULT 7

AAV30142
 ID AAV30142 standard; Protein; 799 AA.

XX
 AC AAV30142;

XX
 DT 26-OCT-1999 (first entry)

```

DE Amino acid sequence of a raffinose synthase protein.
XX
XX Raffinose synthase; plant; broad bean; sucrose; raffinose.
XX
XX
XX
XX
XX
XX Jp11215984-A.
XX
XX 10-AUG-1999.
XX
XX
XX 12-DEC-1997; 97JP-0342899.
XX
XX 28-NOV-1997; 97JP-0329006.
XX 18-DEC-1996; 96JP-0338673.
XX
XX
XX (SUMO ) SUMITOMO CHEM CO LTD.
XX
XX WPI: 1999-511112/43.
XX DR
XX N-PSDB: AA210001.
XX
XX
XX New raffinose synthase gene - is prepared from a plant material
XX
XX
XX Claim 5; Page 19-21; 40pp; Japanese.
XX
XX
XX The present sequence represents a raffinose synthase protein. The
XX sequence is isolated from plant material of broad beans. The
XX protein forms raffinose by complexing alpha(1 to 6)- D-galactosyl
XX hydroxyl group of the 6C of D-glucose residue in sucrose molecules.
XX
XX
XX Sequence 799 AA;

Query Match 81.8%; Score 148; DB 20; Length 799;
Best Local Similarity 80.0%; Pred. No. 5.2e-13;
Matches 24; Conservative 2; Mismatches 4; Indels 0; Gaps 0.

QY 1 FGMCTWDAFYLTVHPGVIEGVRIHVDGSC 30
    |||||
DB 238 fgvctwdaftyikvhpkywgyksiltgdc 267

RESULT 8
AAW57888
ID AAW57888 standard; Protein; 587 AA.
XX
XX
XX AAW57888;
AC
XX
XX 23-SEP-1998 (first entry)
DT
XX
XX Japanese artichoke raffinose synthetase.
DE
XX
XX Raffinose synthetase; metabolism modification; food additive;
XX gastrointestinal flora; Japanese artichoke.
XX
XX Stachys sieboldii.
XX
XX
XX EP849359-A2.
XX
XX
XX 24-JUN-1998.
XX
XX
XX 18-DEC-1997; 97EP-0122417.
XX
XX 18-DEC-1996; 96JP-0338673.
XX
XX (SUMO ) SUMITOMO CHEM CO LTD.
XX
XX
XX Oeda K, Wantanabe E;
XX
XX
XX WPI: 1998-324670/29.
XX DR
XX N-PSDB: AAV40802.
XX
XX New nucleic acid molecule encoding plant raffinose synthetase -
XX capable of producing raffinose, used as food additives with
XX

```

Query Match	Best Local Similarity	80.7%;	Score 146;	DB 19;	Length 587;
Matches	24;	Conservative	2;	Mismatches	4;
				Indels	0;
				Gaps	0;
QY	1	FGWCTWDAFYLVHPQGVIEGVRLVDGCG	30		
DB	107	fgwctwdaftylnvphgymegvgivdggc	136		
RESULT	9				
ID	AAV30144				
	AAV30144 standard; Protein; 587 AA.				
AC	AAV30144;				
DT	26-OCT-1999 (first entry)				
DE	Amino acid sequence of a raffinose synthase protein.				
KW	Raffinose synthase; plant; sucrose; raffinose.				
OS	Stachys sieboldii.				
PN	JPI1215984-A.				
PD	10-AUG-1999.				
PF	12-DEC-1997; 97JP-0342899.				
PR	28-NOV-1997; 97JP-0329006.				
PR	18-DEC-1996; 96JP-0338673.				
XX	(SUMO) SUMITOMO CHEM CO LTD.				
XX	WPI; 1999-511112/43.				
XX	N-PSDB; AA210003.				
PT	New raffinose synthase gene - is prepared from a plant material				
XX	Claim 12; Page 30-31; 40pp; Japanese.				
XX	The present sequence represents a raffinose synthase protein. The				
CC	sequence is isolated from plant material. The protein forms raffinose				
CC	by complexing alpha(1 to 6)- D-galactosyl hydroxyl group of the 6C				
CC	of D-glucose residue in sucrose molecules.				
XX	Sequence	587 AA;			
XX	Sequence	587 AA;			

RESULT 10
AAB98659
ID AAB98659 standard; protein; 780 AA.
XX
AC AAB98659;
XX
DT 17-AUG-2001 (first entry)
XX
DE Soybean protein: SEQ ID 1.
XX
KW Mutant; mutant; raffinose synthase; raffinose oligosaccharide reduction;
XX plant; soybean.
XX
OS Glycine max.
XX
PN JP2001078783-A.
XX
PD 27-MAR-2001.
XX
PF 03-JUL-2000; 2000JP-0200571.
XX
PR 09-JUL-1999; 99JP-0196036.
XX
PA (SUMO) SUMITOMO CHEM CO LTD.
XX
DR MPI: 2001-313373/33.
DR N-PSDB: AAH27438.
XX
XX
PT Novel mutant protein of raffinose synthase is useful for reducing the
PT raffinose oligosaccharide content in a plant body -
XX
PS Disclosure: Page 18-20; 30pp; Japanese.
XX
CC The present invention relates to a mutant protein of raffinose synthase
CC in which at least one aromatic amino acid present at the position of
CC about 1-7 amino acids from the N-terminus is deleted or replaced. The
CC mutant protein can be used for reducing the raffinose oligosaccharide
CC content in a plant body. The present protein from soybean, was used in
CC the present invention.
CC
SQ Sequence 780 AA;

Query Match 80.7%; Score 146; DB 22; Length 780;
Best Local Similarity 80.0%; Pred. No. 1e-12;
Matches 24; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 FGWCTWDAFYLTVPQGVIEGVRLVVGCC 30
DB 220 fgwctwdafyltkvhpqgvegvkglvegcc 249
|||||
RESULT 11
AAM57887
ID AAM57887 standard; Protein; 781 AA.
XX
AC AAM57887;
XX
DT 23-SEP-1998 (first entry)
XX
DE Soybean raffinose synthetase.
XX
KW Raffinose synthetase; metabolism modification; food additive;
KW gastrointestinal flora; soybean.
XX
OS Glycine max.
XX
PN EP849359-A2.
XX
PD 24-JUN-1998.

PF 18-DEC-1997; 97EP-0122417.
XX
XX 18-DEC-1996; 96JP-0338673.
XX
PA (SUMO) SUMITOMO CHEM CO LTD.
XX
XX Oeda K, Wantanabe E;
XX
DR MPI: 1998-324670/29.
DR N-PSDB: AAY40801.
XX
XX
PT New nucleic acid molecule encoding plant raffinose synthetase -
PT capable of producing raffinose, used as food additives with
PT beneficial effects on gastrointestinal flora
XX
PS Claim 1; Page 31-34; 44pp; English.
XX
XX
CC This sequence represents the soybean raffinose synthetase of the
CC invention. The raffinose synthetase is capable of producing raffinose by
CC combining a D-galactosyl group through an alpha (1-6) bond with a
CC hydroxyl group attached to the carbon atom at position 6 of a D-glucose
CC residue in a sucrose molecule. The DNA can be used to modify metabolism
CC of a host organism by introducing into the host organism or its cell so
CC that the content of the raffinose family oligosaccharides in the host
CC organism or cell is changed. Raffinose oligosaccharides are useful as
CC food additives with beneficial effects on the gastrointestinal flora.
XX
SQ Sequence 781 AA;

Query Match 80.7%; Score 146; DB 19; Length 781;
Best Local Similarity 80.0%; Pred. No. 1e-12;
Matches 24; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 FGWCTWDAFYLTVPQGVIEGVRLVVGCC 30
DB 220 fgwctwdafyltkvhpqgvegvkglvegcc 249
|||||
RESULT 12
AAY30143
ID AAY30143 standard; Protein; 781 AA.
XX
AC AAY30143;
XX
DT 26-OCT-1999 (first entry)
XX
DE Amino acid sequence of a raffinose synthase protein.
XX
KW Raffinose synthase; plant; sucrose; raffinose.
XX
OS Glycine max.
XX
PN JP11215984-A.
XX
PD 10-AUG-1999.
XX
PF 12-DEC-1997; 97JP-0342899.
XX
PR 28-NOV-1997; 97JP-0329006.
PR 18-DEC-1996; 96JP-0338673.
XX
PA (SUMO) SUMITOMO CHEM CO LTD.
XX
XX
DR MPI: 1999-511112/43.
DR N-PSDB: AAZ10002.
XX
XX
PT New raffinose synthase gene - is prepared from a plant material
PS Claim 8; Page 25-27; 40pp; Japanese.
XX
XX
CC The present sequence represents a raffinose synthase protein. The
CC sequence is isolated from plant material. The protein forms raffinose

CC by complexing alpha(1 to 6)-D-galactosyl hydroxyl group of the 6C
CC of D-glucose residue in sucrose molecules.
XX
SQ Sequence 781 AA;

Query Match 80.7%; Score 146; DB 20; Length 781;
Best Local Similarity 80.0%; Pred. No. 1e-12;
Matches 24; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 FGWCTWDAFYLTVHPQGVIEGVRHLVDGCG 30
|||||
Db 220 fgwctwdaFYLTkvhpsgywgvkglveggc 249

RESULT 13

AAB49400
ID AAB49400 standard; Protein: 781 AA.

AC AAB49400;

DT 07-MAR-2001 (first entry)

DE Soybean raffinose synthase.

XX Plant promoter; transgenic plant; desired property.

XX Glycine max.

XX EP1048733-A2.

XX 02-NOV-2000.

XX 27-APR-2000; 2000EP-0108962.

XX 30-APR-1999; 99JP-0124527.

XX 01-SEP-1999; 99JP-0247211.

XX (SUMO) SUMITOMO CHEM CO LTD.

XX Ishige F, Watanabe E, Oeda K;

XX WPI; 2001-104537/12.

XX N-PSDB; AAC89523.

XX New soybean plant promoters useful for generating transgenic plants

XX with desired properties -

XX Example 6; Page 24-27; 36pp; English.

XX The present invention provides novel plant promoters which can be used in

XX the production of transgenic plants which express genes with desired

XX properties.

XX Sequence 781 AA;

XX Query Match 80.7%; Score 146; DB 22; Length 781;

XX Best Local Similarity 80.0%; Pred. No. 1e-12;

XX Matches 24; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 FGWCTWDAFYLTVHPQGVIEGVRHLVDGCG 30
|||||
Db 220 fgwctwdaFYLTkvhpsgywgvkglveggc 249

RESULT 14

AAY32075
ID AAY32075 standard; Protein: 572 AA.

AC AAY32075;

DT 17-JAN-2000 (first entry)

XX Rapeseed raffinose synthase.
XX
XX Raffinose synthase; rapeseed; transgenic plant.
XX
XX Brassica napus.

Key Location/Qualifiers
FH Misc-difference 129 /note= "encoded by GCV"
FT
FT Misc-difference 132 /note= "encoded by GCV"
FT
FT Misc-difference 133 /note= "encoded by GCV"
FT
FT Misc-difference 143 /note= "encoded by CCS"
FT
FT Misc-difference 144 /note= "encoded by TCR"
FT
FT Misc-difference 148 /note= "encoded by CGR"

EP953643-A2.

03-NOV-1999.

27-APR-1999; 99EP-0107430.

XX 30-APR-1998; 98JP-0120550.

XX 30-APR-1998; 98JP-0120551.

XX 04-DEC-1998; 98JP-0345590.

XX 10-DEC-1998; 98JP-0351246.

XX (SUMO) SUMITOMO CHEM CO LTD.

XX Watanabe E, Oeda K;

XX WPI; 1999-593144/51.

XX N-PSDB; AA220210.

XX New sense and antisense genes, useful for altering the level of

XX raffinose in food plants -

XX Claim 27; Page 36-38; 55pp; English.

XX This sequence represents rapeseed raffinose synthase, a protein

XX that can bind a D-galactosyl group through an alpha(1-6) bond to the

XX hydroxy group attached to the carbon atom at the 6-position of the

XX D-glucose residue in a sucrose molecule to form raffinose. CDNA

XX (see AA220210) encoding the enzyme was isolated from rapeseed cv.

XX Westar leaf CDNA by PCR. Probes or primers generated from plant

XX other raffinose synthase genes (see AA220207-10) may be used to obtain

XX (claimed). These genes may be used to control the levels of

XX raffinose produced in plants. Antisense genes can be used to knock

XX out existing gene activity, and sense genes to increase the level

XX of gene activity. The resulting transgenic plants may be used as a

XX food source to alter the growing conditions for gut enterobacteria,

XX providing general health advantages.

XX Sequence 572 AA;

Query Match 79.6%; Score 144; DB 20; Length 572;
Best Local Similarity 80.0%; Pred. No. 1.3e-12;
Matches 24; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 FGWCTWDAFYLTVHPQGVIEGVRHLVDGCG 30
|||||
Db 13 fgwctwdaFYLTkvhpsgywgvkglveggc 42

RESULT 15
AAY32074
ID AAY32074 standard; Protein: 777 AA.

```

AC  AAY32074;
XX
DT  17-JAN-2000 (first entry)
XX
DE  Mustard raffinose synthase.
XX
KW  Raffinose synthase; mustard; transgenic plant.
XX
OS  Brassica juncea.
XX
FH  Key Location/Qualifiers
FT  Misc-difference 210
FT  /note= "encoded by ACR"
XX
PN  EP953643-A2.
XX
PD  03-NOV-1999.
XX
PF  27-APR-1999; 99EP-0107430.
XX
PR  30-APR-1998; 98JP-0120550.
PR  30-APR-1998; 98JP-0120551.
PR  04-DEC-1998; 98JP-0345590.
PR  10-DEC-1998; 98JP-0351246.
XX
PA  (SUMO ) SUMITOMO CHEM CO LTD.
XX
PI  Watanabe E, Oeda K;
XX
DR  WPI; 1999-593144/51.
DR  N-PSDB; AA220209.
XX
PT  New sense and antisense genes, useful for altering the level of
PT  raffinose in food plants .
XX
PS  Claim 26; Page 29-31; 55pp; English.
XX
CC  This sequence represents mustard raffinose synthase, a protein
CC  that can bind a D-galactosyl group through an alpha(1-6) bond to the
CC  hydroxy group attached to the carbon atom at the 6-position of the
CC  D-glucose residue in a sucrose molecule to form raffinose. CDNA
CC  (see AA220209) encoding the enzyme was isolated from mustard
CC  (Brassica juncea) leaf cDNA by PCR. Probes or primers generated from
CC  plant raffinose synthase genes (see AA220207-10) may be used to obtain
CC  other raffinose synthase genes by labeled detection or amplification
CC  (claimed). These genes may be used to control the levels of
CC  raffinose produced in plants. Antisense genes can be used to knock
CC  out existing gene activity, and sense genes to increase the level
CC  of gene activity. The resulting transgenic plants may be used as a
CC  food source to alter the growing conditions for gut enterobacteria,
CC  providing general health advantages.
XX
SQ  Sequence 777 AA;

```

```

Query Match 79.6%; Score 144; DB 20; Length 777;
Best Local Similarity 80.0%; Pred. No. 1.9e-12;
Matches 24; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
QY 1 FGMCTWDATYLVHPQGVIEGVRLVLDGCG 30
DB 218 fgwctwdatyltvpdgvhkvkclvdggc 247

```

Search completed: November 30, 2001, 09:42:56
Job time: 155 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 30, 2001, 09:40:21 ; Search time 39.46 Seconds.

(without alignments)
57.913 Million cell updates/sec

Title: US-08-846-234-1

Perfect score: 181

Sequence: 1 FGWCTMDAFYLVHPGVIEGVRLVDGCG 30

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	111	61.3	765	2	probable imbibition
2	108	59.7	357	2	probable seed imbi
3	108	59.7	757	2	S1p1 protein - Bar
4	108	59.7	773	2	imbibition protein
5	106	58.6	807	2	hypothetical prote
6	82	45.3	1170	2	protein F20N2.14
7	72	39.8	371	2	T01717
8	64.5	35.6	374	2	E82168
9	60	33.1	137	2	146711
10	60	33.1	146	1	S00185
11	60	33.1	154	2	A31674
12	60	33.1	154	2	C37360
13	60	33.1	157	1	GKROB
14	60	33.1	210	2	A32398
15	60	33.1	451	2	T24018
16	59	32.6	164	2	S31622
17	53.5	29.6	333	2	A85620
18	53.5	29.6	333	2	G64833
19	53	29.3	189	2	A48834
20	53	29.3	557	2	E85744
21	52.5	29.0	265	2	E82964
22	52	28.7	237	2	B86420
23	52	28.7	619	2	D86509
24	52	28.7	619	2	D81556
25	52	28.7	619	2	A72114
26	52	28.7	1462	1	B36182
27	51	28.2	106	1	ERAD31
28	51	28.2	106	2	S52799
29	51	28.2	106	2	S52809

30	51	28.2	1039	2	F71427	hypothetical prote
31	50	27.6	381	2	T23250	hypothetical prote
32	49.5	27.3	310	2	S38592	prolyl aminopeptid
33	49	27.1	79	2	S35456	GAM1 protein - Pla
34	49	27.1	191	1	H64744	yead protein - Esc
35	49	27.1	191	2	B85505	probable phosphata
36	49	27.1	227	2	S76494	hypothetical prote
37	49	27.1	589	2	T29897	hypothetical prote
38	49	27.1	1296	2	S55511	hypothetical prote
39	48.5	26.8	302	2	B35961	valine--tRNA ligas
40	48.5	26.8	515	1	A32931	hypothetical 21k p
41	48.5	26.8	522	2	S75491	glucosylceramidase
42	48.5	26.8	536	1	EUNHGC	hypothetical prote
43	48.5	26.8	861	2	T00434	glucosylceramidase
44	48	26.5	141	2	H83229	probable kinase h
45	48	26.5	207	2	JC5940	conserved hypothet
						fibroblast growth

ALIGNMENTS

RESULT 1
S45033
probable imbibition protein - wild cabbage
C:Species: Brassica oleracea (wild cabbage)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Sep-1997
C:Accession: S45033
R:Fujikura, Y.; Karsen, C.K.
submitted to the EMBL Data Library, May 1994
A:Description: Cauliflower cDNA encoding a putative imbibition protein.
A:Reference number: S45033
A:Accession: S45033
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-765 <FUJ>
A:Cross-references: EMBL:X79330; NID:g488786; PID:g488787

Query Match 61.3%; Score 111; DB 2; Length 765;
Best Local Similarity 65.5%; Pred. No. 1.7e-07;
Matches 19; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

OY 1 FGWCTMDAFYLVHPGVIEGVRLVDGCG 29
DB 198 FGWCTMDAFYLVHPGVIEGVRLVDGCG 226

RESULT 2
T09530
probable seed imbibition protein - chickpea (fragment)
C:Species: Cicer arietinum (chickpea, garbanzo)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 20-Jun-2000
C:Accession: T09530
R:Cervantes, E.
submitted to the EMBL Data Library, February 1996
A:Reference number: Z16718
A:Accession: T09530
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-357 <CER>
A:Cross-references: EMBL:X95875; NID:g1212811
A:Experimental source: germinating seed
A:Genetics:
A:Gene: s1p

Query Match 59.7%; Score 108; DB 2; Length 357;
Best Local Similarity 62.1%; Pred. No. 1.9e-07;
Matches 18; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

OY 1 FGWCTMDAFYLVHPGVIEGVRLVDGCG 29
DB 31 FGWCTMDAFYLVHPGVIEGVRLVDGCG 59

RESULT 3
S27762
Slp1 protein - barley
C:Species: Hordeum vulgare (barley)
C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 09-Sep-1997
C:Accession: S27762
R:Heck, G.R.; Dorsett, C.; Ho, T.H.
submitted to the EMBL Data Library, February 1992
A:Description: Cloning and characterization of a gene, Slp1, associated with seed imbibition
A:Reference number: S27762
A:Accession: S27762
A:Molecule type: DNA
A:Residues: 1-757 <HEC>
A:Cross-references: EMBL:M7475; NID:g167099; PID:g167100
C:Genetics:
A:Introns: 64/2; 151/1; 181/2; 197/3; 489/3; 557/2

Query Match 59.7%; Score 108; DB 2; Length 757;
Best Local Similarity 62.1%; Pred. No. 4.3e-07;
Matches 18; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

OY 1 FGCTWDAAFLTLVHPQGVIEGRHLVDGG 29
|||||
db 205 FGCTWDAAFLTYDVTALGKQGLRSLAEEG 233
|||||

RESULT 4
T46188
Imbibition protein homolog - Arabidopsis thaliana
N:Alternate names: protein TBH10.120
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C:Accession: T46188
R:Benes, V.; Rechmann, S.; Borkova, D.; Anstorge, W.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.
submitted to the Protein Sequence Database, January 2000
A:Reference number: Z23014
A:Accession: T46188
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-773 <BEN>
A:Cross-references: EMBL:AL133248
A:Experimental source: cultivar Columbia; BAC clone TBH10
C:Genetics:
A:Map position: 3
A:Insertions: 64/2; 146/1; 176/2; 192/3; 223/2; 259/2; 300/2; 484/3; 507/2; 552/2; 625/1; 6
A:Note: TBH10.120

```

Query Match Similarity      59.7%;      Score 108;  DB 2;  Length 773;
Best Local Similarity      62.1%;      Pred. No. 4.4e-07;
Matches      18;  Conservative      4;  Mismatches      7;  Indels      0;  Gaps      0;

QY      1  FGCWCTDAFYLTIVHPOGVIEGVRLHLYDGC 29
          |||||  |||  |  :||  |||  :||  ||
DB      200  FGCWCTDAFYLTIVTAEGVDEGLKLSSEGG 228

```

```

RESULT      5
C850025
hypothetical protein AT4g01970 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001
C:Accession: C850025
R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring
Nature 402, 769-777, 1999
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A:Reference number: AB5001; MUID:20083488
A:Accession: C850025
A:Status: preliminary
A:Molecule type: DNA

```

A;Residues: 1-807 <STO>
A;Cross-references: GB_NC_001268; NTD:g7268581; PIDN:CAB80690.1; GSPDB:GN001400
C;Genetics:
A;Gene: AT4g01970
A;Map position: 4

[illegible]

RESULT 6
C69599
protein F20N2.14 [imported] - Arabidopsis thaliana
C.Species: Arabidopsis thaliana (mouse-ear cress)
C.Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C.Accession: C69599
R.Rthologis: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, N.F.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, T.; Hansen, N.F.; Hughes, B.; Huizart, L.
Nature 408, 816-820, 2000
A.Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marzita, R.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A.Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo, Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A.Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A.Reference number: A86141; MUID:21016719
A.Accession: C69599
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-1170 <STO>
A.Cross-references: GB:AE005173; NID:98778496; PIDN:AAE79504.1; GSPDB:GN00141
C.Genetics:
A.Gene: F20N2.14
A.Map position: 1

Query Match	Score 82;	DB 2;	Length 1170;
Best Local Similarity	51.9%;	Pred. No. 0.0029;	
Matches 14;	Conservative	3;	Mismatches 10; Indels 0; Gaps 0;
QY	1	FGMCTMDAFLYLVHPQGVLEGVRHLVD	27
Db	203	FGMCTMDAFLYLVTAADVAQGLESCND	229

```

RESULT
7
T01717
hypothetical protein A_IG002N01.5 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 24-Mar-1999
C:Accession: T01717
R:Schaeff, P.; Maggi, L.
submitted to the EMBL Data Library, June 1997
A:Description: The sequence of A. thaliana IG002N01.
A:Reference number: Z14407
A:Accession: T01717
A:Status: translated from GR/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-371 <SCH>
A:Cross-references: EMBL:AF007269; NID:g2191126; PID:g2191156
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 4
A:Intons: 5/2; 25/3; 103/3; 126/1; 174/2; 240/1; 284/1; 295/1; 333/2; 349/2
A:Note: A_IG002N01.5

```

Query Match 39.8%; Score 72; DB 2; Length 371;
 C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C:Accession: S00185
 C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
 R:Stimpson, R.J.; Moritz, R.L.; Lloyd, C.J.; Fabrl, L.J.; Nice, E.C.; Rubira, M.R.; Bu
 FES Lett. 224, 128-132, 1987
 A:Title: Primary structure of ovine pituitary basic fibroblast growth factor.
 A:Reference number: S00185; MUID:88055577
 A:Accession: S00185

OY 1 FGCTWDAFYLVHPGVIEGVRLVD 13
 Db 111 FGCTWDAFYRVK 123

RESULT 8
 E82168
 probable alpha-1,6-galactosidase VC1690 [imported] - Vibrio cholerae (strain N16961 sero
 C:Species: Vibrio cholerae
 C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
 C:Accession: E82168
 R:Heidelberger, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;
 chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, F
 l, R.R.; Melanos, J.J.; Venter, J.C.; Fraser, C.M.
 Nature 406, 477-483, 2000
 A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
 A:Reference number: A82035; MUID:20406833
 A:Accession: E82168
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-374 <HEI>
 A:Cross-references: GB:AE004247; GB:AE003852; NID:g9656204; PIDN:AAF94840.1; GSPDB:GN001
 A:Experimental source: serogroup O1; strain N16961; biotype El Tor
 C:Genetics:
 A:Gene: VC1690
 A:Map position: 1

Query Match 35.6%; Score 64.5; DB 2; Length 374;
 Best Local Similarity 38.5%; Pred. No. 0.24;
 Matches 15; Conservative 4; Mismatches 9; Indels 11; Gaps 2;

OY 2 GWCWDAFYLVHPGVIEGV-----RH-----LVNDCG 29
 Db 197 GWCWDAFYLVHPGVIEGV-----RH-----LVNDCG 235

RESULT 9
 I46711
 fibroblast growth factor - rabbit (fragment)
 C:Species: Oryctolagus cuniculus (domestic rabbit)
 C>Date: 14-Feb-1997 #sequence_revision 14-Feb-1997 #text_change 16-Jul-1999
 C:Accession: I46711
 R:Winkler, J.A.; Friesel, R.; Alberts, G.F.; Janat, M.F.; Llau, G.
 Am. J. Pathol. 143, 518-527, 1993
 A:Title: Elevated expression of basic fibroblast growth factor in an immortalized rabbit
 A:Reference number: I46711; MUID:93343209
 A:Accession: I46711
 A:Status: preliminary
 A:Molecule type: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-137 <WIN>
 A:Cross-references: GB:LI2034; NID:g165014; PIDN:AAA31248.1; PID:g165015
 C:Superfamily: fibroblast growth factor

Query Match 33.1%; Score 60; DB 2; Length 137;
 Best Local Similarity 40.0%; Pred. No. 0.35;
 Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

OY 3 WCTWDAFYLVHPGVIEGVRLVD 27
 Db 24 YCKNGFFLRHPDGRVGVREKSD 48

RESULT 10
 S00185
 basic fibroblast growth factor - sheep
 N:Alternate names: prostatropin

C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
 C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C:Accession: S00185
 C:Species: Rattus norvegicus (Norway rat)
 N:Alternate names: bFGF
 C:Date: 21-May-1990 #sequence_revision 21-May-1990 #text_change 16-Jul-1999
 C:Accession: A31674; S00876; S24309
 R:Shimasaki, S.; Emoto, N.; Koba, A.; Mercado, M.; Shibata, F.; Cooksey, K.; Baird, A
 Biochem. Biophys. Res. Commun. 157, 256-263, 1988
 A:Title: Complementary DNA cloning and sequencing of rat ovarian basic fibroblast gro
 A:Reference number: A31674; MUID:89061721
 A:Accession: A31674

A:Molecule type: mRNA
 A:Residues: 1-154 <SHI>
 A:Cross-references: GB:M22427; NID:g204285; PIDN:AAA1210.1; PID:g204286
 R:Kurokawa, T.; Seno, M.; Igarashi, K.
 Nucleic Acids Res. 16, 5201, 1988
 A:Title: Nucleotide sequence of rat basic fibroblast growth factor cDNA.
 A:Reference number: S00876; MUID:88262516
 A:Accession: S00876
 A:Molecule type: mRNA
 A:Residues: 1-154 <RUR>
 A:Cross-references: EMBL:X07285; NID:g56203; PIDN:CAA30265.1; PID:g56204
 R:El-Husseini, A.E.D.; Paterson, J.A.; Myal, Y.; Shiu, R.P.C.
 Biochim. Biophys. Acta 1131, 314-316, 1992
 A:Title: PCR detection of the rat brain basic fibroblast growth factor (bFGF) mRNA co
 A:Reference number: S24309; MUID:92339546
 A:Accession: S24309
 A:Status: preliminary
 A:Molecule type: translation not shown
 A:Molecule type: mRNA
 A:Residues: 35-154 <SHI>
 A:Cross-references: EMBL:X61697; NID:g56143; PIDN:CAA43863.1; PID:g56144
 C:Superfamily: fibroblast growth factor
 C:Keywords: growth factor
 F:1-9/Domain: signal sequence #status predicted <SIG>
 F:10-154/Product: basic fibroblast growth factor #status predicted <MAT>

Query Match 33.1%; Score 60; DB 1; Length 146;
 Best Local Similarity 40.0%; Pred. No. 0.37;
 Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

OY 3 WCTWDAFYLVHPGVIEGVRLVD 27
 Db 24 YCKNGFFLRHPDGRVGVREKSD 48

RESULT 11
 A31674
 basic fibroblast growth factor precursor - rat

A:Alternate names: bFGF
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 21-May-1990 #sequence_revision 21-May-1990 #text_change 16-Jul-1999
 C:Accession: A31674; S00876; S24309
 R:Shimasaki, S.; Emoto, N.; Koba, A.; Mercado, M.; Shibata, F.; Cooksey, K.; Baird, A
 Biochem. Biophys. Res. Commun. 157, 256-263, 1988
 A:Title: Complementary DNA cloning and sequencing of rat ovarian basic fibroblast gro
 A:Reference number: A31674; MUID:89061721
 A:Accession: A31674
 A:Molecule type: mRNA
 A:Residues: 1-154 <SHI>
 A:Cross-references: GB:M22427; NID:g204285; PIDN:AAA1210.1; PID:g204286
 R:Kurokawa, T.; Seno, M.; Igarashi, K.
 Nucleic Acids Res. 16, 5201, 1988
 A:Title: Nucleotide sequence of rat basic fibroblast growth factor cDNA.
 A:Reference number: S00876; MUID:88262516
 A:Accession: S00876
 A:Molecule type: mRNA
 A:Residues: 1-154 <RUR>
 A:Cross-references: EMBL:X07285; NID:g56203; PIDN:CAA30265.1; PID:g56204
 R:El-Husseini, A.E.D.; Paterson, J.A.; Myal, Y.; Shiu, R.P.C.
 Biochim. Biophys. Acta 1131, 314-316, 1992
 A:Title: PCR detection of the rat brain basic fibroblast growth factor (bFGF) mRNA co
 A:Reference number: S24309; MUID:92339546
 A:Accession: S24309
 A:Status: preliminary
 A:Molecule type: translation not shown
 A:Molecule type: mRNA
 A:Residues: 35-154 <SHI>
 A:Cross-references: EMBL:X61697; NID:g56143; PIDN:CAA43863.1; PID:g56144
 C:Superfamily: fibroblast growth factor
 C:Keywords: growth factor
 F:1-9/Domain: signal sequence #status predicted <SIG>
 F:10-154/Product: basic fibroblast growth factor #status predicted <MAT>

Query Match 33.1%; Score 60; DB 2; Length 154;
 Best Local Similarity 40.0%; Pred. No. 0.4;
 Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

OY 3 WCTWDAFYLVHPGVIEGVRLVD 27
 Db 32 YCKNGFFLRHPDGRVGVREKSD 56

RESULT 12

```

C37360
basic fibroblast growth factor - mouse
C:Species: Mus musculus (house mouse)
C>Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 16-Jul-1999
C:Accession: C37360
R:Hebert, J.M.; Basillio, C.; Goldfarb, M.; Haub, O.; Martin, G.R.
Dev. Biol. 138, 454-463, 1990
A:Title: Isolation of cDNAs encoding four mouse FGF family members and characterization
A:Reference number: A37360; MUID:90201563
A:Accession: C37360
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-154 <HEB>
A:Cross-references: GB:M30644; NID:g193296; PIDN:AAA37621.1; PID:g309239
C:Superfamily: fibroblast growth factor

Query Match      33.1%  Score 60; DB 2; Length 154;
Best Local Similarity 40.0%  Pred. No. 0.4;
Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

OY      3  WCTWDAPFLTVHPGVIGVRHLVD 27
       1 1 1 1 1 1 1 1 1 1 1 1 1
Db      32  YCKNGCFRLRIHPDGRVDGVREKSD 56

RESULT 13
GKBOB
basic fibroblast growth factor precursor - bovine (fragment)
N:Alternate names: bFGF, kidney-derived growth factor; prostatiotropin
C:Species: Bos primigenius taurus (cattle)
C:Date: 13-Aug-1986 #sequence_revision 02-Jun-1995 #text_change 24-Nov-1999
C:Accession: A24663; A32878; A33784; A61550; A60310; A61094; A01386; A60316; A22
R:Abraham, J.A.; Mergia, A.; Whang, J.L.; Tumolo, A.; Friedland, J.; Hjertild, K.A.; Gosp
Science 233, 545-548, 1986
A:Title: Nucleotide sequence of a bovine clone encoding the angiogenic protein, basic fi
A:Reference number: A94290; MUID:86261806
A:Accession: A24663
A:Molecule type: mRNA
A:Residues: 3-157 <ABR>
A:Cross-references: GB:M3440; NID:g163049; PIDN:AAA30518.1; PID:g163050
A:Experimental source: pituitary gland
R:Abraham, J.A.; Whang, J.L.; Tumolo, A.; Mergia, A.; Fiddes, J.C.
Cold Spring Harb. Symp. Quant. Biol. 51, 657-668, 1986
A:Title: Human basic fibroblast growth factor: nucleotide sequence, genomic organization
A:Reference number: A90924; MUID:87217066
A:Accession: A32878
A:Molecule type: mRNA
A:Residues: 3-157 <AB2>
R:Miller, P.G.; Li, Y.S.; Hoffman, R.M.; Kodner, C.M.; Siegel, N.R.; Deuel, T.F.
Biochem. Biophys. Res. Commun. 165, 1096-1103, 1989
A:Title: A novel 17 kD heparin-binding growth factor (HBGF-8) in bovine uterus: purifica
A:Reference number: A33784; MUID:90121211
A:Accession: A33784
A:Molecule type: protein
A:Residues: 1-14 <MLT>
A:Note: demonstration of a possible alternative initiator or splice junction
R:Bertolini, J.; Hearn, M.T.W.
Mol. Cell. Endocrinol. 51, 187-199, 1987
A:Title: Isolation, characterization and tissue localisation of an N-terminal-truncated
A:Reference number: A61550; MUID:87247652
A:Accession: A61550
A:Molecule type: protein
A:Residues: 16-35 <BER>
R:Ueno, N.; Baird, A.; Esch, F.; Ling, N.; Guillemin, R.
Mol. Cell. Endocrinol. 49, 189-194, 1987
A:Title: Isolation and partial characterization of basic fibroblast growth factor from h
A:Reference number: A61551; MUID:87162856
A:Accession: A61551
A:Molecule type: protein
A:Residues: 27-35, 'X', 37-41 <BE3>
A:Experimental source: testes
A:Note: this form appears to be identical to the renal form

```

```

R:Ueno, N.; Baird, A.; Esch, F.; Shimasaki, S.; Ling, N.; Guillemin, R.
Regul. Pept. 16, 135-145, 1986
A:Title: Purification and partial characterization of a mitogenic factor from bovine
A:Reference number: A60310; MUID:87119165
A:Accession: A60310
A:Molecule type: protein
A:Residues: 23-35, 'X', 37-42 <UEN>
A:Experimental source: liver
R:Ueno, N.; Baird, A.; Esch, F.; Ling, N.; Guillemin, R.
Biochem. Biophys. Res. Commun. 138, 580-588, 1986
A:Title: Isolation of an amino terminal extended form of basic fibroblast growth fact
A:Reference number: A24819; MUID:86295737
A:Contents: annotation
A:Note: the amino end of this form was blocked; the peptide composition matched what
R:Gospodarowicz, D.; Baird, A.; Cheng, J.; Lui, G.M.; Esch, F.; Bohlen, P.
Endocrinology 118, 82-90, 1986
A:Title: Isolation of fibroblast growth factor from bovine adrenal gland: physicochem
A:Reference number: A61094; MUID:86081530
A:Accession: A61094
A:Molecule type: protein
A:Residues: 12-25, 27-35, 'X', 37-40 <GOS>
A:Experimental source: adrenal gland
R:Esch, F.; Baird, A.; Ling, N.; Ueno, N.; Hill, F.; Denoroy, L.; Klepper, R.; Gospod
Proc. Natl. Acad. Sci. U.S.A. 82, 6507-6511, 1985
A:Title: Primary structure of bovine pituitary basic fibroblast growth factor (FGF) a
A:Reference number: A01386; MUID:86016731
A:Accession: A01386
A:Molecule type: protein
A:Residues: 12-157 <ESC>
A:Experimental source: pituitary gland
R:Baird, A.; Esch, F.; Bohlen, P.; Ling, N.; Gospodarowicz, D.
Regul. Pept. 12, 201-213, 1985
A:Title: Isolation and partial characterization of an endothelial cell growth factor
A:Reference number: A60316; MUID:86095426
A:Accession: A60316
A:Molecule type: protein
A:Residues: 27-35, 'X', 37-43 <BAI>
A:Experimental source: kidney
R:Bohlen, P.; Baird, A.; Esch, F.; Ling, N.; Gospodarowicz, D.
Proc. Natl. Acad. Sci. U.S.A. 81, 5364-5368, 1984
A:Title: Isolation and partial molecular characterization of pituitary fibroblast gro
A:Reference number: A22054; MUID:84298139
A:Accession: A22054
A:Molecule type: protein
A:Residues: 12-26 <BOH>
C:Comment: The acidic and basic fibroblast growth factors are the major endothelial-c
all types in vitro (although bFGF is 30-100 times more potent than aFGF in stimulatn
C:Superfamily: fibroblast growth factor
C:Keywords: alternative splicing; angiogenesis; blocked amino end; growth factor; hep
F:1-157/Product: basic fibroblast growth factor, uterine form #status predicted <MAT1
F:14-157/Product: basic fibroblast growth factor, pituitary alpha form #status experim
F:12-157/Product: basic fibroblast growth factor, pituitary gamma form #status experim
F:16-157/Product: basic fibroblast growth factor, pituitary short form #status predic
F:23-157/Product: basic fibroblast growth factor, hepatic form #status experimental <
F:27-157/Product: basic fibroblast growth factor, renal form #status experimental <MA
F:29-33, 118-121/Region: heparin binding #status predicted
F:4/Modified site: blocked amino end (Ala) (in mature form pituitary gamma) (probably

```

```

Query Match      33.1%  Score 60; DB 1; Length 157;
Best Local Similarity 40.0%  Pred. No. 0.4;
Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

OY      3  WCTWDAPFLTVHPGVIGVRHLVD 27
       1 1 1 1 1 1 1 1 1 1 1 1 1
Db      35  YCKNGCFRLRIHPDGRVDGVREKSD 59

RESULT 14
A32398
basic fibroblast growth factor precursor, 22-5k form - human
N:Alternate names: bFGF, fibroblast growth factor 2; prostatic growth factor; prostat

```


RESULT 15

T24018

hypothetical protein R07B7.11 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000

C/Accession: T24018

R:Harris, B.

submitted to the EMBL Data Library, July 1996

A/Reference number: Z19830

A/Accession: T24018

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-451 <WIL>

A/Cross-references: EMBL:Z75955; PIDN:CAB00120.1; GSPDB:GM00023; CESP:R07B7.11

A/Experimental source: clone R07B7

C/Genetics:

A/Gene: CESP:R07B7.11

A/Map position: 5

A/Introns: 61/1; 107/3; 317/3; 351/3

C/Superfamily: alpha-galactosidase

Query Match 33.1%; Score 60; DB 2; Length 451;

Best Local Similarity 36.8%; Pred. No. 1.2;

Matches 14; Conservative 3; Mismatches 11; Indels 10; Gaps 2;

QY 2 GWCCTWDAFYLTV---HPGVI-----EGVRHLVDGG 29

DB 27 GWMSTWTFYCEIDCVKHPGTCINQGLYKDMADQLVSSG 64

Search completed: November 30, 2001, 09:41:40
Job time: 79 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 30, 2001, 09:40:21 : Search time 23.6 seconds
(without alignments)

46,608 Million cell updates/sec

Title: US-08-846-234-1
Perfect score: 181
Sequence: 1 FGWCTWDAFYLTVHPGVIEGVRLVVDGCG 30

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 100059 seqs, 36664827 residues
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	60	33.1	137	1 FGF2_RABIT	P48799 oryctolagus
2	60	33.1	154	1 FGF2_MOUSE	P15655 mus musculus
3	60	33.1	154	1 FGF2_RAT	P13109 rattus norv
4	60	33.1	155	1 FGF2_BOVIN	P03669 bos taurus
5	60	33.1	155	1 FGF2_HUMAN	P09038 homo sapien
6	60	33.1	155	1 FGF2_SHEEP	P20003 ovis aries
7	59	32.6	156	1 FGF2_MONDO	P48798 monodelphis
8	53.5	29.6	319	1 SSUA_ECOLI	P75853 escherichia
9	53	29.3	158	1 FGF2_CHICK	P48800 gallus gall
10	52	28.7	1462	1 PRPE_DROME	P16620 drosophila
11	51	28.2	106	1 P312_ADE03	P11319 human adeno
12	49.5	27.3	310	1 P1P_NERGO	P42786 neisseria g
13	49	27.1	191	1 YABD_ECOLI	P31546 escherichia
14	49	27.1	227	1 LIPB_SYNY3	P74519 synchocyst
15	48.5	26.8	302	1 VIAL_PSESS	P18205 pseudomonas
16	48.5	26.8	302	1 GLCM_MOUSE	P17439 mus musculu
17	48.5	26.8	536	1 GLCM_HUMAN	P04062 homo sapien
18	48	26.5	207	1 FGF6_HUMAN	O43320 homo sapien
19	48	26.5	207	1 FGF6_RAT	O54769 rattus norv
20	48	26.5	366	1 CHSH_IPOTF	P48402 ipomoea tril
21	48	26.5	388	1 DXR_ZYMO	O9X5F2 zymomonas m
22	48	26.5	468	1 LACG_STRMU	P50978 streptococc
23	48	26.5	470	1 LACG_STRAU	P11175 staphylococ
24	47.5	26.2	369	1 MAGA_HUMAN	P43363 homo sapien
25	47.5	26.2	500	1 C72X_ARATH	O911P3 arabidopsis
26	47.5	26.2	510	1 C72X_ARATH	O911P3 arabidopsis
27	47	26.0	247	1 FGF2_MOUSE	O92915 homo sapien
28	47	26.0	247	1 FGF2_MOUSE	P70379 mus musculu
29	47	26.0	341	1 GRMA_COMPX	P07385 compox virtu
30	47	26.0	406	1 DXR_MYCLE	O9CBU3 mycobacteri
31	47	26.0	413	1 DXR_MYCTU	O10798 mycobacteri
32	47	26.0	548	1 HEMI_YEAST	P09950 saccharomyc
33	47	26.0	686	1 VILL_HUMAN	O15195 homo sapien

34	46.5	25.7	299	1 RFBJ_SALTY	P22716 salmonella
35	46.5	25.7	498	1 NDD_ALCXX	P94212 alcaigenes
36	46.5	25.7	521	1 OPN4_MOUSE	O9GXXZ mus musculu
37	46.5	25.7	582	1 HEMO_OPSTA	P43090 opasanus tau
38	46	25.4	132	1 VE4_HPV2A	P25483 human papil
39	46	25.4	232	1 SPIB_VACCC	P20842 vaccinia vi
40	46	25.4	352	1 CHSC_IPOPU	P48399 ipomoea pur
41	46	25.4	379	1 DXR_CHLMU	O9PKW8 chlamydia m
42	46	25.4	380	1 DXR_AQUAE	O66722 aquifex aco
43	46	25.4	396	1 CHSA_IPOPU	P48397 ipomoea pur
44	45.5	25.1	105	1 E312_ADE12	P36706 human adeno
45	45.5	25.1	369	1 P311_HUMAN	P21128 homo sapien

ALIGNMENTS

RESULT 1
ID FGF2_RABIT STANDARD: PRT: 137 AA.
AC P48799:
DT 01-FEB-1996 (rel. 33, Created)
DT 01-FEB-1996 (rel. 33, Last sequence update)
DT 01-FEB-1996 (rel. 33, Last annotation update)
DE HEPARIN-BINDING GROWTH FACTOR 2 (HBGF-2) (BASIC FIBROBLAST GROWTH
DE FACTOR) (BFGF) (PROSTATROPIN) (FRAGMENT).
GN FGF2
OS Oryctolagus cuniculus (Rabbit).
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=NEW ZEALAND WHITE; TISSUE=Smooth muscle;
RC MEDLINE=93343209; PubMed=8342599;
RA Winkles J.A., Friesel R., Alberts G.F., Janet M.F., Lau G.;
RT "Elevated expression of basic fibroblast growth factor in an
RT immortalized rabbit smooth muscle cell line."
RL Am. J. Pathol. 143:518-527(1993).
CC - FUNCTION: THE HEPARIN-BINDING GROWTH FACTORS ARE ANGIOGENIC AGENTS
CC IN VIVO AND ARE POTENT MITOGENS FOR A VARIETY OF CELL TYPES IN
CC VITRO. THERE ARE DIFFERENCES IN THE TISSUE DISTRIBUTION AND
CC CONCENTRATION OF THESE 2 GROWTH FACTORS.
CC - SUBUNIT: MONOMER.
CC - MISCELLANEOUS: THIS PROTEIN BINDS HEPARIN MORE STRONGLY THAN DOES
CC AFGF.
CC - SIMILARITY: BELONGS TO THE HEPARIN-BINDING GROWTH FACTORS FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
DR EMBL: L12034; AAA31248.1; -.
DR HSSP: P09038; 1BFF.
DR InterPro: IPR002209; HBGF_FGF.
DR Pfam: PF00167; FGF_1.
DR ProDom: PD000831; HBGF_FGF; 1.
DR SMART: SM00442; FGF_1; FGF_1.
DR PROSITE: PS00247; HBGF_FGF; 1.
KW Growth factor; Mitogen; Vascularization; Heparin-binding.
FT BINDING 18 22 HEPARIN (POTENTIAL).
FT BINDING 107 110 HEPARIN (POTENTIAL).
FT NON_TER 137 137
SQ SEQUENCE 137 AA; 15418 MW; 0D9EE457BB8BE8C51 CRC64;

Query Match 33.1%; Score 60; DB 1; Length 137;
Best local Similarity 40.0%; Pred. No. 0.16;
Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

FT	TURN	45	46
FT	STRAND	49	52
FT	TURN	55	56
FT	HELIX	58	60
FT	STRAND	62	68
FT	TURN	69	70
FT	STRAND	71	76
FT	TURN	77	80
FT	STRAND	81	85
FT	TURN	87	88
FT	STRAND	91	94
FT	STRAND	99	101
FT	-HELIX	103	107
FT	STRAND	109	110
FT	TURN	113	117
FT	STRAND	121	122
FT	TURN	124	124
FT	STRAND	127	127
FT	STRAND	129	130
FT	TURN	133	133
FT	STRAND	136	138
FT	HELIX	141	142
FT	TURN	144	146
FT	HELIX	148	151
FT	STRAND	155 AA:	17250 MM:
50	SEQUENCE		BE6CE70FA6107129 CRC64;

RA [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=67162468; PubMed=2435575;
 RA Kurokawa T., Sasada R., Iwane M., Igarashi K.;
 RT "Cloning and expression of cDNA encoding human basic fibroblast
 RT growth factor.";
 RL FEBS Lett. 213:189-194(1987).
 RN [5]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=69184522; PubMed=2538817;
 RA Prats H., Kaghad M., Prats A.C., Klagsbrun M., Lelias J.M.,
 RA Llaunzu P., Chalou P., Tauber J.P., Amalric F., Smith J.A.,
 RA Caput D.;
 RT "High molecular mass forms of basic fibroblast growth factor are
 RT initiated by alternative CUG codons.";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:1836-1840(1989).
 RN [6]
 RN SEQUENCE OF 10-35.
 RX MEDLINE=86275260; PubMed=37322516;
 RA Gautschi P., Frater-Schroeder M., Boehlen P.;
 RA "Partial molecular characterization of endothelial cell mitogens from
 RA human brain: acidic and basic fibroblast growth factors.";
 RL FEBS Lett. 204:203-207(1986).
 RN [7]
 RN SEQUENCE OF 10-39.
 RX MEDLINE=86186784; PubMed=3964255;
 RA Gimenez-Gallego G., Conn G., Hatcher V.B., Thomas K.A.;
 RA "Human brain-derived acidic and basic fibroblast growth factors:
 RA amino terminal sequences and specific mitogenic activities.";
 RL Biochem. Biophys. Res. Commun. 135:541-548(1986).
 RN [8]
 RN SEQUENCE OF 2-22.
 RX MEDLINE=87156686; PubMed=2435284;
 RA Story M.T., Esch F., Shimaseki S., Sasse J., Jacobs S.C., Lawson R.K.,
 RA "Amino-terminal sequence of a large form of basic fibroblast growth
 RA factor isolated from human benign prostatic hyperplastic tissue.";
 RL Biochem. Biophys. Res. Commun. 142:702-709(1987).
 RN [9]
 RN X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
 RP MEDLINE=91195367; PubMed=1707342;
 RX Eriksson A.E., Couzens L.S., Weaver L.H., Matthews B.W.;
 RA "Three-dimensional structure of human basic fibroblast growth
 RA factor.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:3441-3445(1991).
 RN [10]
 RN X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).
 RP MEDLINE=94004464; PubMed=7691311;
 RX Eriksson A.E., Couzens L.S., Matthews B.W.;
 RA "Refinement of the structure of human basic fibroblast growth factor
 RA at 1.6-A resolution and analysis of presumed heparin binding sites by
 RA selenate substitution.";
 RL Protein Sci. 2:1274-1284(1993).
 RN [11]
 RN X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
 RP MEDLINE=91195368; PubMed=1849557;
 RX Zhang J., Couzens L.S., Barr P.J., Sprang S.R.;
 RA "Three-dimensional structure of human basic fibroblast growth factor,
 RA a structural homolog of interleukin 1 beta.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:3446-3451(1991).
 RN [12]
 RN X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).
 RP MEDLINE=92121151; PubMed=1769967;
 RX Ago H., Kitagawa Y., Fujishima A., Matsuura Y., Katsube Y.;
 RA "Crystal structure of basic fibroblast growth factor at 1.6-A
 RA resolution.";
 RL J. Biochem. 110:360-363(1991).
 RN [13]
 RN X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
 RP MEDLINE=91095983; PubMed=1702556;
 RX Zhu X., Komiya H., Chirino A., Faham S., Fox G.M., Arakawa T.,
 RA Hsu B.T., Rees D.C.;
 RT "Three-dimensional structures of acidic and basic fibroblast growth
 RT factors."


```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AJ237695; CAB40390.1; -
DR EMBL: AE000195; AAC74022.1; ALT_INIT.
DR EMBL: D90732; BAA35691.1; ALT_INIT.
DR Ecogene: EG13707; ssua.
DR InterPro: IPR001638; SBP_bac_3.
DR SMART: SM00062; Bbpf.1.
DR Transprot: Periplasmic; Signal: Complete proteome.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 319 PUTATIVE ALIPHATIC SULFONATES BINDING
FT PROTEIN.
SQ SEQUENCE 319 AA; 34557 MW; BACB0711F38D53D0 CRC64;

Query Match 29.6%; Score 53.5; DB 1; Length 319;
Best Local Similarity 46.2%; Pred. No. 2.9;
Matches 12; Conservative 1; Mismatches 10; Indels 3; Gaps 1;

OY 3 WCTMDAFYLVHPGVIEGVRHLVD 28
DB 180 WAIWDPPYSNALLQG---GVRVKKDG 202

RESULT 9
FGF2_CHICK STANDARD: PRT; 158 AA.
AC P48800;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE HEPARIN-BINDING GROWTH FACTOR 2 PRECURSOR (HBGF-2) (BASIC FIBROBLAST
DE GROWTH FACTOR) (HBGF).
GN FGF2 OR FGF-2.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauromorpha; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_Taxid=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93246053; PubMed=7683281;
RA Bojta A.Z., Zeller R., Meljers C.;
RT "Expression of alternatively spliced hbfg first coding exons and
RT anti-sense mRNAs during chicken embryogenesis.";
RL Dev. Biol. 157:110-118(1993).
CC -1- FUNCTION: THE HEPARIN-BINDING GROWTH FACTORS ARE ANGIOGENIC AGENTS
CC IN VIVO AND ARE POTENT MITOGENS FOR A VARIETY OF CELL TYPES IN
CC VITRO. THERE ARE DIFFERENCES IN THE TISSUE DISTRIBUTION AND
CC CONCENTRATION OF THESE 2 GROWTH FACTORS.
CC -1- SUBUNIT: MONOMER.
CC -1- MISCELLANEOUS: THIS PROTEIN BINDS HEPARIN MORE STRONGLY THAN DOES
CC AFGF.
CC -1- SIMILARITY: BELONGS TO THE HEPARIN-BINDING GROWTH FACTORS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M95707; AAA48617.1; -
DR HSSP: P09038; IBEF.
DR InterPro: IPR002209; HBGF_FGF.

```

```

DR InterPro: IPR002348; ILL_HBGF.
DR Pfam: PF00167; FGF.1.
DR PRINTS: PR00262; ILL_HBGF.
DR PRINTS: PR00263; HBGF_FGF.
DR PRINTS: PD000831; HBGF_FGF.1.
DR PRODOM: PD000442; FGF.1.
DR SMART: SM00442; FGF.1.
DR PROSITE: PS00247; HBGF_FGF.1.
DR Growth factor; Mitogen; Vascularization; Heparin-binding.
KM PROPEP 1 12 BY SIMILARITY.
FT CHAIN 13 158 HEPARIN-BINDING GROWTH FACTOR 2.
FT BINDING 30 34 HEPARIN (POTENTIAL).
FT BINDING 119 122 HEPARIN (POTENTIAL).
SQ SEQUENCE 158 AA; 17374 MW; 7B69B684C17F1816 CRC64;

Query Match 29.3%; Score 53; DB 1; Length 158;
Best Local Similarity 36.0%; Pred. No. 1.7;
Matches 9; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

OY 3 WCTMDAFYLVHPGVIEGVRHLVD 27
DB 36 YCKNGGFRLINPDGRVDGVRKSD 60

RESULT 10
PPP6_DROME STANDARD: PRT; 1462 AA.
ID PPP6_DROME
AC P16620;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROTEIN-TYROSINE PHOSPHATASE DPPP PRECURSOR (EC 3.1.3.48) (PROTEIN-
DE TYROSINE PHOSPHATE PHOSPHORYLASE).
GN PPP69D OR DPPP.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90046860; PubMed=2554325;
RA Streuli M., Krueger N.X., Tsai A.Y.M., Saito H.;
RT "A family of receptor-linked protein tyrosine phosphatases in humans
RT and Drosophila.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:8698-8702(1989).
CC -1- FUNCTION: IT IS POSSIBLE THAT DPPP IS A CELL ADHESION RECEPTOR.
CC -1- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =
CC PROTEIN TYROSINE + ORTHOPHOSPHATE.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M27699; AAA28842.1; -
DR PIR: B36182; B36182.
DR HSSP: P18052; IYFO.
DR FLYBASE: FBgn0014007; Ptp69D.
DR InterPro: IPR001777; FN_III.
DR InterPro: IPR003006; IG_MHC.
DR InterPro: IPR003598; IG_C2.
DR InterPro: IPR003600; IG_Like.
DR InterPro: IPR00387; TYR_phosphatase.
DR InterPro: IPR00242; TYR_prot_phphatase.
DR Pfam: PF00041; fn3; 3.

```



```

DR Pfam: PF00561; abhydrolase; 1.
DR PRINTS; PR00793; PROAMINOPEPTIDASE.
KW Hydrolase; Aminopeptidase.
FT ACT_SITE 107 107 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 260 260 BY SIMILARITY.
FT ACT_SITE 287 287 PROTON DONOR (BY SIMILARITY).
SQ SEQUENCE 310 AA; 34791 MW; D0E7AEBA908A1AE0 CRC64;

Query Match 27.3%; Score 49.5; DB 1; Length 310;
Best Local Similarity 23.7%; Pred. No. 10;
Matches 9; Conservative 7; Mismatches 11; Indels 11; Gaps 1;

OY 3 WCTWDAFLVTHPQGVIEGV-----RHLYDGG 29
DB 195 WADMESYLRFEPGVDEDAVATLARIENHYFVNGG 232

RESULT 13
YAEI_ECOLI STANDARD; PRT; 191 AA.
AC P31546;
DT 01-JUL-1993 (rel. 26, Created)
DT 01-JUL-1993 (rel. 26, Last sequence update)
DT 20-AUG-2001 (rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN YAEI.
GN YAEI OR B0200 OR Z0212 OR ECS0202.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RL Miyamoto K.;
RN Submitted (APR-1993) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX BLattner F.R., Plunkett G., III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / W3110;
RX Takemoto K., Mori H., Murayama N., Kataoka K., Yano M., Itoh T.,
RA Yamamoto Y., Inokuchi H., Miki T., Hatada E., Fukuda R., Ichihara S.,
RA Mizuno T., Makino K., Nakata A., Yura T., Sempel G., Mizobuchi K.;
RT "Systematic sequencing of the Escherichia coli genome: analysis of the
RT 4.0 - 6.0 min (189,987 - 281,416bp) region.";
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
[4]
RP SEQUENCE FROM N.A.
RC Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,
RA Davis K., Federspiel N., Hyman R., Kaiman S., Komp C., Kurdi O.,
RA Lashkari D., Lew H., Lin D., Nemeth A., Oefner P., Roberts D.,
RA Davis R.W.;
RT Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
[5]
RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / RIMD 0509952;
RA Ohnishi M., Murata T., Nakayama K., Kuhara S., Hattori M.,
RA Kurokawa K., Yasunaga T., Makino K., Shinagawa H., Hayashi T.;
RT "Development of primer sets for direct sequence determination of all
RT the ribosomal operons of Escherichia coli.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
[6]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN-O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE-21074935; PubMed-11206551;
RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoudis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
[7]
RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / RIMD 0509952;
RX MEDLINE-21156231; PubMed-11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
CC -1- SIMILARITY: STRONG, TO H.INFLUENZAE HI0621.1.
CC -1- SIMILARITY: TO M.JANNASCHII MJ0015.
CC -1- SIMILARITY: TO THE N-TERMINAL PART OF E. COLI H1SB, THE HISTIDINOL
CC PHOSPHATE PHOSPHATASE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch).
CC -----
DR EMBL; D15061; BAA03661.1; ALT_INIT.
DR EMBL; AE000129; AAC73311.1; -.
DR EMBL; D83536; BAA77877.1; -.
DR EMBL; U70214; AAB08628.1; -.
DR EMBL; AB035926; BAA93568.1; -.
DR EMBL; AE005195; AAG54502.1; -.
DR EMBL; AP002550; BAB3625.1; -.
DR Ecocycle; EG11736; yaei.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 191 AA; 21294 MW; E7014B34A23128FA CRC64;

Query Match 27.1%; Score 49; DB 1; Length 191;
Best Local Similarity 28.6%; Pred. No. 7; 6;
Matches 10; Conservative 3; Mismatches 12; Indels 10; Gaps 1;

OY 3 WCTWDAFLVTHPQGVIEGVRLVD 27
DB 74 WMDWSLADRDVLDGIYCPHHPGSGVEEFGVCD 108

RESULT 14
LIPB_SYNT3 STANDARD; PRT; 227 AA.
AC P74519;
DT 15-JUL-1998 (rel. 36, Created)
DT 15-JUL-1998 (rel. 36, Last sequence update)
DT 20-AUG-2001 (rel. 40, Last annotation update)
DE LIPOATE-PROTEIN LIGASE B (EC 6.-.-.-) (LIPOATE BIOSYNTHESIS PROTEIN
DE B).
GN LIPB OR SLR0994.
OS Synecocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synecocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-97061201; PubMed-8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirosewa M., Sugliara M., Saito M., Saito S., Kimura T.,

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 30, 2001, 09:40:21 ; Search time 67.45 Seconds
(without alignments)
65.058 Million cell updates/sec

Title: US-08-846-234-1
Perfect score: 181
Sequence: 1 FGWCTMDAFYLVHPGVIEGVRLVYDGGC 30

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP-REMBL-17:*
2: sp-archaea:*
3: sp-bacteria:*
4: sp-fungi:*
5: sp-human:*
6: sp-invertebrate:*
7: sp-mammal:*
8: sp-mhc:*
9: sp-organellar:*
10: sp-plant:*
11: sp-rodent:*
12: sp-virus:*
13: sp-vertebrate:*
14: sp-unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	181	100.0	784	10	Q92T62 Cucumis sat
2	144	79.6	783	10	Q9FND9 Arabidopsis
3	117	64.6	857	10	Q9SBZ0 Phaseolus a
4	111	61.3	765	10	Q43408 Brassica o
5	108	59.7	357	10	Q39466 Cicer ariet
6	108	59.7	757	10	Q40077 Hordeum vul
7	108	59.7	773	10	Q9SCM1 Arabidopsis
8	106	58.6	807	10	Q9SXT4 Arabidopsis
9	101	55.8	779	10	Q9M4M7 Persea amer
10	82	45.3	1170	10	Q9LFZ7 Arabidopsis
11	72	39.8	371	10	Q04607 Arabidopsis
12	64.5	35.6	374	2	Q9KRF2 Cucurbita
13	60	33.1	114	4	Q00527 Homo sapien
14	60	33.1	114	4	Q16443 Homo sapien
15	60	33.1	130	6	Q77767 Canis fami
16	60	33.1	196	4	P78443 Homo sapien
17	60	33.1	451	5	Q21801 Caenorhabd
18	57	31.5	170	11	Q60487 Canis fami
19	56.5	31.2	221	5	Q9GPR2 Crassostrea

20	54	29.8	390	10	Q43812 Pelargonium
21	54	29.8	489	10	Q43811 Pelargonium
22	54	29.8	489	10	P93772 Pelargonium
23	53	29.3	557	2	Q9KEU2 Bacillus th
24	52.5	29.0	265	2	P72162 Pseudomonas
25	52.5	29.0	265	2	Q9HTB8 Pseudomonas
26	52	28.7	237	10	Q9C7N3 Chlamydia p
27	52	28.7	619	2	Q9E935 Chlamydia p
28	52	28.7	619	2	Q9K228 Chlamydia p
29	52	28.7	1428	2	Q9VU03 Drosophila
30	51	28.2	106	12	Q83115 Mastadenov
31	51	28.2	106	12	Q83124 Mastadenov
32	51	28.2	283	10	Q9FVM2 Arabidopsis
33	51	28.2	417	10	Q9FXT4 Oryza sativ
34	51	28.2	989	10	Q23723 Arabidopsis
35	51	28.2	1039	10	Q23461 Arabidopsis
36	50	27.6	284	2	Q9L413 Salimella
37	50	27.6	381	5	Q9XUV4 Caenorhabd
38	50	27.6	386	2	Q9AJD7 Klebsiella
39	50	27.6	417	3	Q02402 Mortierella
40	49.5	27.3	352	5	Q9GXY1 Leishmania
41	49	27.1	80	5	Q26176 Plasmodium
42	49	27.1	80	5	Q27313 Plasmodium
43	49	27.1	180	2	Q9KXL5 Streptomyces
44	49	27.1	293	2	Q9L866 Aspergillus
45	49	27.1	321	4	Q9H1X0 Homo sapien

ALIGNMENTS

RESULT	ID	Q92T62	PRELIMINARY:	PRT:	784 AA.
AC	Q92T62;				
DT	01-MAY-1999 (TREMBLrel. 10, Created)				
DT	01-MAY-1999 (TREMBLrel. 10, Last sequence update)				
DT	01-MAY-2000 (TREMBLrel. 13, Last annotation update)				
DE	RAEFTNOSE SYNTHASE (EC 2.4.1.82).				
GN	RFS.				
OS	Cucumis sativus (Cucumber).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;				
OC	eurosid 1; Cucurbitales; Cucurbitaceae; Cucumis.				
OX	NCBI_TaxID=3659;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=LEAF;				
RA	Ohsuni C., Nozaki J., Kida T.;				
RL	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AF073744; AAD02832.1; -				
KW	Transferase; Glycosyltransferase.				
SQ	SEQUENCE 784 AA; 86919 MW; 3B06A491F0908933 CRC64;				

Query Match	100.0%;	Score 181;	DB 10;	Length 784;
Best Local Similarity	100.0%;	Pred. No. 9e-17;		
Matches 30;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1	FGWCTMDAFYLVHPGVIEGVRLVYDGGC 30	
DB	215	FGWCTMDAFYLVHPGVIEGVRLVYDGGC 244	

RESULT	2	
Q9FND9		PRELIMINARY:
AC	Q9FND9;	PRT: 783 AA.
DT	01-MAR-2001 (TREMBLrel. 16, Created)	
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)	
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)	
DE	RAEFTNOSE SYNTHASE PROTEIN.	
OS	Arabidopsis thaliana (Mouse-ear cress).	

```
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=980659011; PubMed=9405937;
RA Kotani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. II.
RT Sequence features of the regions of 1,044,062 bp covered by thirteen
RT physically assigned P1 clones.";
RL DNA Res. 4:291-300(1997).
DR EMBL: AB006702; BABI1595.1; -.
SQ SEQUENCE 783 AA; 86237 MW; 3C37D1D7871888AC CRC64;

Query Match 79.6%; Score 144; DB 10; Length 783;
Best Local Similarity 80.0%; Pred. No. 1.3e-11;
Matches 24; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 1 FGCTWDAFYLTVPQGVIEGRHLVDGCG 30
DB 224 FGCTWDAFYLTVPDGVHKGCVKCLVDGCG 253

RESULT 3
O9SB20 PRELIMINARY; PRT; 857 AA.
AC O9SB20;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE GALACTINOL-RAFFINOSE GALACTOSYLTRANSFERASE (EC 2.4.1.67).
OS Phaseolus angularis (Adzuki bean) (Vigna angularis).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids I; Fabales; Fabaceae; Papilionoideae; Vigna.
OX NCBI_TaxID=3914;
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=20117502; PubMed=10652123;
RA Peterbauer T., Mucha J., Mayer U., Popp M., Gloessel J., Richter A.;
RT "Synthesis of stachyose in seeds of adzuki bean (Vigna angularis).
RT Molecular cloning and functional expression of stachyose synthase from
RT adzuki bean.";
RL Plant J. 20:509-518(1999).
DR EMBL: Y19024; CAB64363.1; -.
KM Transferase; Glycosyltransferase.
SQ SEQUENCE 857 AA; 94949 MW; 85248C4B81165679 CRC64;

Query Match 64.6%; Score 117; DB 10; Length 857;
Best Local Similarity 65.5%; Pred. No. 8e-08;
Matches 19; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY 1 FGCTWDAFYLTVPQGVIEGRHLVDGCG 29
DB 221 FGCTWDAFYLTVPVGVHGLKDFSEGG 249

RESULT 4
O43408 PRELIMINARY; PRT; 765 AA.
AC O43408;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE PUTATIVE INHIBITION PROTEIN.
OS Brassica oleracea (Cauliflower).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
```

```
OC eucosids II; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID=3712;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=CV. ALPHA BALOMA; TISSUE=ROOT TIPS;
RA Fujikura Y., Karszen C.K.;
RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL: X79330; CAA55893.1; -.
DR Mendel, 30784; Braol.1449;30784.
SQ SEQUENCE 765 AA; 84084 MW; 73398603048E9B58 CRC64;

Query Match 61.3%; Score 111; DB 10; Length 765;
Best Local Similarity 65.5%; Pred. No. 4.8e-07;
Matches 19; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

OY 1 FGCTWDAFYLTVPQGVIEGRHLVDGCG 29
DB 198 FGCTWDAFYTDVTAEGVDEGLSLSEGG 226

RESULT 5
O39466 PRELIMINARY; PRT; 357 AA.
AC O39466;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE SEED INHIBITION PROTEIN (FRAGMENT).
GN SIP.
OS Cicer arietinum (Chickpea) (Garbanzo).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids I; Fabales; Fabaceae; Papilionoideae; Cicer.
OX NCBI_TaxID=3827;
RN
RP SEQUENCE FROM N.A.
RA Cervantes E.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: X95875; CAA65125.1; -.
DR Mendel; 7745; Cicer.1449;7745.
FT NON_TER 357 357
FT SEQUENCE 357 AA; 39465 MW; E8CD30399E1BF3EE CRC64;

Query Match 59.7%; Score 108; DB 10; Length 357;
Best Local Similarity 62.1%; Pred. No. 5.2e-07;
Matches 18; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

OY 1 FGCTWDAFYLTVPQGVIEGRHLVDGCG 29
DB 31 FGCTWDAFYTDVTAEGVDEGLKSLSEGG 59

RESULT 6
O40077 PRELIMINARY; PRT; 757 AA.
AC O40077;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE SEED INHIBITION PROTEIN.
GN SIPL.
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
OC Triticeae; Hordeum.
OX NCBI_TaxID=4513;
RN
RP SEQUENCE FROM N.A.
RA Heck G.R., Dorsett C., Ho T.H.;
RL Submitted (FEB-1992) to the EMBL/GenBank/DBJ databases.
```



```
RP SEQUENCE FROM N.A.
RA Khan S., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Kim C.,
RA Shin P., Altati H., Bel O., Chin C., Chlou J., Choi E., Conn L.,
RA Conway A., Gonzales A., Hansen N.N.F., Howng B., Koo T., Lam B., Lee J.,
RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharasy N., Nguyen M.,
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
RA Ecker J.R.;
RT "Genomic sequence for Arabidopsis thaliana BAC F20N2 from chromosome
RT I.";
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RP Ecker J.R.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RN SEQUENCE FROM N.A.
RP Ecker J.R.;
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RN SEQUENCE FROM N.A.
RP Shin P.P., Buehler E.E., Dunn P.P., Feng J.J., Kim C.C., Li Y.Y.,
RA Walker M.W., Altati H.H., Araujo R.R., Conn L.L., Conway A.A.B.,
RA Gonzalez A.A., Hansen N.N.F., Huizar L.L., Kremenetskaya I.I.,
RA Lenz C.C., Li J.J., Liu S.S., Lueros S.S., Rowley D.D., Schwartz J.J.,
RA Toriumi M.M., Vysotskaya V.V., Yu G.G., Davis R.R.W.,
RA Federspiel N.N.A., Theologis A.A., Ecker J.J.R.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [5]
RN SEQUENCE FROM N.A.
RP Ecker J.R.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [6]
RN SEQUENCE FROM N.A.
RA Cheuk R., Shin P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
RA Khan S., Kim C., Altati H., Bel B., Chin C., Chlou J., Choi E., Lam B.,
RA Conn L., Conway A., Gonzalez A., Hansen N., Howng B., Koo T., Lam B.,
RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharasy N.,
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
RA Theologis A., Ecker J.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC002328; AAF79504.1;
SQ SEQUENCE 1170 AA; 129057 MW; 98B43A04B3F6D44 CRC64;

Query Match 45.3%; Score 82; DB 10; Length 1170;
Best Local Similarity 51.9%; Pred. No. 0.0084;
Matches 14; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

QY 1 FGWCTWDAFYLTVPQVIEGVRLVLD 27
DB 203 FGWCTWDAFYLTVPQVIEGVRLVLD 229

RESULT 11
ID 004607 PRELIMINARY; PRT; 371 AA.
AC 004607;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE SIMILAR TO SEED IMBIBITION PROTEIN.
GN A_LG002N01.5
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Scheet P., Magl L.;
```

```
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Waterston R.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF007269; AAB61043.1;
DR Wendei, 16897; Arath.1449;16897.
SQ SEQUENCE 371 AA; 41564 MW; 9AEX7712D348344F CRC64;

Query Match 39.8%; Score 72; DB 10; Length 371;
Best Local Similarity 84.6%; Pred. No. 0.056;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FGWCTWDAFYLTVP 13
DB 111 FGWCTWDAFYRKV 123

RESULT 12
ID 09KRF2 PRELIMINARY; PRT; 374 AA.
AC 09KRF2;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE ALPHA-1,6-GALACTOSIDASE, PUTATIVE.
GN VCI1690.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; Pubmed=10952301;
RA Heidelberg J.F., Eissen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.J., Hart D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gills S.R., Nelson K.E., Read T.D., Tellein H., Richardson D.,
RA Ermolaeva M.D., Yamathayan J., Bass S., Qin H., Dragol I., Sellers P.,
RA McDonald L., Uterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RA "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RT Nature 406:477-483(2000).
RL EMBL: AE004247; AAF94840.1;
DR TIGR: VCI1690;
KW Complete proteome.
SQ SEQUENCE 374 AA; 42962 MW; 44EF6804239B98B4 CRC64;

Query Match 35.6%; Score 64.5; DB 2; Length 374;
Best Local Similarity 38.5%; Pred. No. 0.62;
Matches 15; Conservative 4; Mismatches 9; Indels 11; Gaps 2;

QY 2 GWCTWDAFYLTVPQVIEGV-----RH-----LVDCG 29
DB 197 GWCSMYAYVAEVTQDIKENVAIIAERHPELEWVLLDDG 235

RESULT 13
ID 000527 PRELIMINARY; PRT; 114 AA.
AC 000527;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1999 (TREMBLrel. 09, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE BASIC FIBROBLAST GROWTH FACTOR (FRAGMENT).
GN FGF-2 OR FGF2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 30, 2001, 09:40:21 ; Search time 34.66 Seconds
(without alignments)
19.478 Million cell updates/sec

Title: US-08-846-234-1

Perfect score: 181
Sequence: 1 FGMCTWDARFLYLVHPQGVIEGVRLVLDGCG 30

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents -AA: *
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep: *
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep: *
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep: *
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep: *
5: /cgn2_6/ptodata/2/1aa/PCYUS.COMB.pep: *
6: /cgn2_6/ptodata/2/1aa/Backfiles1.pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	181	100.0	30	4	US-08-846-234-1
2	181	100.0	784	4	US-08-846-234-5
3	60	33.1	132	1	US-08-100-744-4
4	60	33.1	132	1	US-08-284-784-4
5	60	33.1	132	2	US-08-854-811-4
6	60	33.1	140	5	PCT-US90-06962-1
7	60	33.1	145	1	US-07-830-330-1
8	60	33.1	145	1	US-08-187-780-3
9	60	33.1	145	2	US-08-478-485-3
10	60	33.1	146	2	US-08-231-894A-11
11	60	33.1	146	2	US-08-231-894A-12
12	60	33.1	146	2	US-08-231-894A-13
13	60	33.1	146	2	US-08-231-894A-14
14	60	33.1	146	2	US-08-231-894A-15
15	60	33.1	146	2	US-09-105-678A-49
16	60	33.1	146	4	US-09-421-208-49
17	60	33.1	146	6	5464943-6
18	60	33.1	146	6	5464943-8
19	60	33.1	146	6	5464943-10
20	60	33.1	146	6	5464943-12
21	60	33.1	146	6	5464943-14
22	60	33.1	146	6	5464943-25
23	60	33.1	146	6	5464943-26
24	60	33.1	147	6	5175147-8
25	60	33.1	147	6	5314872-1
26	60	33.1	150	1	US-08-441-629-8
27	60	33.1	150	3	US-08-776-207-8

28	60	33.1	150	5	PCT-US95-09172-8	Sequence 8, Appl1
29	60	33.1	153	3	US-08-325-186-2	Sequence 2, Appl1
30	60	33.1	154	2	US-08-438-439C-24	Sequence 24, Appl1
31	60	33.1	154	3	US-08-325-186-1	Sequence 1, Appl1
32	60	33.1	154	5	PCT-US91-02186-6	Sequence 6, Appl1
33	60	33.1	155	1	US-07-959-369-6	Sequence 6, Appl1
34	60	33.1	155	1	US-07-959-369-7	Sequence 7, Appl1
35	60	33.1	155	1	US-08-023-757-2	Sequence 2, Appl1
36	60	33.1	155	1	US-08-023-757-4	Sequence 4, Appl1
37	60	33.1	155	1	US-07-842-177A-1	Sequence 1, Appl1
38	60	33.1	155	1	US-08-177-502-2	Sequence 2, Appl1
39	60	33.1	155	1	US-08-177-502-4	Sequence 4, Appl1
40	60	33.1	155	1	US-08-439-725A-10	Sequence 10, Appl1
41	60	33.1	155	1	US-08-325-632-1	Sequence 1, Appl1
42	60	33.1	155	1	US-08-462-169B-10	Sequence 10, Appl1
43	60	33.1	155	2	US-08-867-471-10	Sequence 10, Appl1
44	60	33.1	155	2	US-08-438-439C-14	Sequence 14, Appl1
45	60	33.1	155	6	5514566-8	Patent No. 5514566

ALIGNMENTS

RESULT 1
US-08-846-234-1
; Sequence 1, Application US/08846234
; Patent No. 6166292
; GENERAL INFORMATION:
; APPLICANT: OSUMI Chieko
; APPLICANT: NOZAKI Jinshi
; APPLICANT: KIDA Takao
; TITLE OF INVENTION: RAFFINOSE SYNTHASE GENE, METHOD FOR
; TITLE OF INVENTION: PRODUCING RAFFINOSE, AND TRANSGENIC PLANT
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATIER & NEUSTADT, P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,234
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)-413-3000
; TELEFAX: (703)-413-2220
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
; US-08-846-234-1

Query Match 100.0%; Score 181; DB 4; Length 30;
Best Local Similarity 100.0%; Pred. No. 3.9e-20;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FGMCTWDARFLYLVHPQGVIEGVRLVLDGCG 30
|||||
DB 1 FGMCTWDARFLYLVHPQGVIEGVRLVLDGCG 30

```
RESULT 2
US-08-846-234-5
; Sequence 5, Application US/08846234
; Patent No. 6166292
; GENERAL INFORMATION:
; APPLICANT: OSUMI Chieko
; APPLICANT: KIDA Takao
; TITLE OF INVENTION: RAFINOSE SYNTHASE GENE, METHOD FOR
; TITLE OF INVENTION: PRODUCING RAFINOSE, AND TRANSGENIC PLANT
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT, P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,234
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)-413-3000
; TELEFAX: (703)-413-2220
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 784 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-846-234-5

Query Match 100.0%; Score 181; DB 4; Length 784;
Best Local Similarity 100.0%; Pred. NO. 2e-18;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FGMCTMDAFYLVHPGVIGVGRHLVDGCG 30
DB 215 FGMCTMDAFYLVHPGVIGVGRHLVDGCG 244

RESULT 3
US-08-100-744-4
; Sequence 4, Application US/08100744
; Patent No. 5563046
; GENERAL INFORMATION:
; APPLICANT: MASCARENHAS, DESMOND
; APPLICANT: ZHANG, SUNNY
; APPLICANT: OLSON, PAMELA
; APPLICANT: OLSEN, DAVID
; APPLICANT: CARILLO, PEDRO A.
; TITLE OF INVENTION: POLYPEPTIDE FUSIONS TO
; TITLE OF INVENTION: INTERLEUKIN-1-LIKE POLYPEPTIDES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/100,744
FILING DATE: 02-AUG-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: LUTHER, BARBARA J.
REGISTRATION NUMBER: 33,954
REFERENCE/DOCKET NUMBER: 22095-20275.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 813-5600
TELEFAX: (415) 494-0792
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 132 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-100-744-4

Query Match 33.1%; Score 60; DB 1; Length 132;
Best Local Similarity 40.0%; Pred. NO. 0.13;
Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 3 WCTMDAFYLVHPGVIGVGRHLVD 27
DB 10 YCKNGFFLRHPDGRVDGVRKSD 34

RESULT 4
US-08-284-784-4
; Sequence 4, Application US/08284784
; Patent No. 5629172
; GENERAL INFORMATION:
; APPLICANT: MASCARENHAS, DESMOND
; APPLICANT: ZHANG, YANG
; APPLICANT: OLSON, PAMELA S.
; APPLICANT: OLSEN, DAVID R.
; APPLICANT: CARILLO, PEDRO A.
; TITLE OF INVENTION: EXPRESSION OF FUSION POLYPEPTIDES
; TITLE OF INVENTION: TRANSPORTED OUT OF THE CYTOPLASM WITHOUT LEADER SEQUENCES
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/284,784
; FILING DATE: 02-AUG-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: PARK, FREDIE K.
; REGISTRATION NUMBER: 35,636
; REFERENCE/DOCKET NUMBER: 22095-20275.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 813-5600
; TELEFAX: (415) 494-0792
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 132 amino acids
```



```

: APPLICATION NUMBER: US/08/478,485
: FILING DATE: Concurrently Herewith
: CLASSIFICATION: 424
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: 08/187,780
: FILING DATE: January 25, 1994
: APPLICATION NUMBER: 07/901,705
: FILING DATE: June 22, 1992
: APPLICATION NUMBER: 07/806,771
: FILING DATE: December 6, 1991
: APPLICATION NUMBER: 07/177,506
: FILING DATE: April 4, 1988
: APPLICATION NUMBER: 07/062,925
: FILING DATE: June 16, 1987
: ATTORNEY/AGENT INFORMATION:
: NAME: Joseph R. Robinson
: REGISTRATION NUMBER: 33,448
: REFERENCE/DOCKET NUMBER: 5986/13586-US6
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 527-7700
: TELEFAX: (212) 753-6237
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 145
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE:
: DESCRIPTION: Protein
: FEATURE:
: NAME/KEY:
: LOCATION:
: IDENTIFICATION METHOD:
: OTHER INFORMATION: This sequence,
: OTHER INFORMATION: corresponding to bovine basic fibroblast
: OTHER INFORMATION: growth factor, can be found in Table 1,
: OTHER INFORMATION: page 9, lines 9, 14, and 19, in the
: OTHER INFORMATION: application, as filed.
: PUBLICATION INFORMATION:
: AUTHORS:
: TITLE:
: JOURNAL:
: VOLUME:
: ISSUE:
: PAGES:
: DATE:
: DOCUMENT NUMBER:
: FILING DATE:
: PUBLICATION DATE:
: RELEVANT RESIDUES IN SEQ ID NO: 1-145
: US-08-478-485-3
:
: Query Match 33.1%; Score 60; DB 2; Length 145;
: Best Local Similarity 40.0%; Pred. No. 0.15;
: Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0.
:
: QY 3 WCTMDAFYLVHPGVGICVGRHLVD 27
: | | | | | | | | | |
: Db 24 YCKNGFELRIHPDGRVDCVGRKSD 48
:
: RESULT 10
: US-08-231-894A-11
: Sequence 11, Application US/08231894A
: Patent No. 5851990
: GENERAL INFORMATION:
: APPLICANT: FUJISHIMA, AKIRA
: APPLICANT: FUKUDA, TSUNEHIKO
: TITLE OF INVENTION: BRGF MUTEIN AND ITS PRODUCTION
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS

```

```

ADDRESSSEE:  & CUSHMAN
STREET:  130 WATER STREET
CITY:  BOSTON
STATE:  MASSACHUSETTS
COUNTRY:  US
ZIP:  02109
COMPUTER READABLE FORM:
MEDIUM TYPE:  Diskette
COMPUTER:  IBM Compatible
OPERATING SYSTEM:  DOS
SOFTWARE:  FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER:  US/08/231,894A
FILING DATE:  22-APR-1994
CLASSIFICATION:  435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:  US 07/873907
FILING DATE:  24-APR-1992
CLASSIFICATION:  435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:  JP 097655-1991
FILING DATE:  26-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER:  JP 066381-1992
FILING DATE:  24-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME:  RESNICK, DAVID S.
REGISTRATION NUMBER:  34235
REFERENCE/DOCKET NUMBER:  41769-FWC
TELECOMMUNICATION INFORMATION:
TELEPHONE:  (617) 523-3400
TELEFAX:  (617) 523-6440
INFORMATION FOR SEQ ID NO:  11:
SEQUENCE CHARACTERISTICS:
LENGTH:  146 amino acids
TYPE:  amino acid
STRANDEDNESS:  single
TOPOLOGY:  linear
MOLECULE TYPE:  peptide
HYPOTHETICAL:  NO
ANTI-SENSE:  NO
FRAGMENT TYPE:  Internal
ORIGINAL SOURCE:
US-08-231-894A-11

Query Match 33.1%; Score 60; DB 2; Length 146;
Best Local Similarity 40.0%; Pred. No. 0.15;
Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY      3 WCTWDAYFLTVHPQGVICGRHLVD 27
      :|:|:|:|:|:|:|:|:|
Db      24 YCKNGCFRLRHPOGRVDGVREKSD 48

RESULT 11
US-08-231-894A-12
: Sequence 12, Application US/08231894A
: Patent No. 5851990
: GENERAL INFORMATION:
: APPLICANT: FUJISHIMA, AKIRA
: APPLICANT: FUKUDA, TSONEHIKO
: TITLE OF INVENTION: BGF MUTEIN AND ITS PRODUCTION
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: DAVID G. CONLIN, DIKE, BRONSTEIN, ROBERTS
: ADDRESSEE:  & CUSHMAN
: STREET:  130 WATER STREET
: CITY:  BOSTON
: STATE:  MASSACHUSETTS
: COUNTRY:  US
: ZIP:  02109
: COMPUTER READABLE FORM:

```

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/231,894A
FILING DATE: 22-APR-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/873907
FILING DATE: 24-APR-1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 097655-1991
FILING DATE: 26-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 066381-1992
FILING DATE: 24-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: RESNICK, DAVID S.
REGISTRATION NUMBER: 34235
REFERENCE/DOCKET NUMBER: 41769-FWC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 523-3400
TELEFAX: (617) 523-6440
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 146 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
US-08-231-894A-12

Query Match 33.1%; Score 60; DB 2; Length 146;
Best Local Similarity 40.0%; Pred. No. 0.15;
Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;
QY 3 WCTWDATYLVHPQVIEGVRLVD 27
Db 24 YCKNGGFLRIHPDGRVDGVREKSD 48
RESULT 12
US-08-231-894A-13
Sequence 13, Application US/08231894A
Patent No. 5851990
GENERAL INFORMATION:
APPLICANT: FUJISHIMA, AKIRA
APPLICANT: FUKUDA, TSUNEHICO
TITLE OF INVENTION: BGF MUTIN AND ITS PRODUCTION
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: DAVID G. CONLIN, DIKE, BRONSTEIN, ROBERTS
ADDRESSEE: 6 CUSHMAN
STREET: 130 WATER STREET
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: US
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/231,894A
FILING DATE: 22-APR-1994

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/873907
FILING DATE: 24-APR-1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 097655-1991
FILING DATE: 26-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 066381-1992
FILING DATE: 24-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: RESNICK, DAVID S.
REGISTRATION NUMBER: 34235
REFERENCE/DOCKET NUMBER: 41769-FWC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 523-3400
TELEFAX: (617) 523-6440
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 146 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
US-08-231-894A-13

Query Match 33.1%; Score 60; DB 2; Length 146;
Best Local Similarity 40.0%; Pred. No. 0.15;
Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 3 WCTWDATYLVHPQVIEGVRLVD 27
Db 24 YCKNGGFLRIHPDGRVDGVREKSD 48

RESULT 13
US-08-231-894A-14
Sequence 14, Application US/08231894A
Patent No. 5851990
GENERAL INFORMATION:
APPLICANT: FUJISHIMA, AKIRA
APPLICANT: FUKUDA, TSUNEHICO
TITLE OF INVENTION: BGF MUTIN AND ITS PRODUCTION
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: DAVID G. CONLIN, DIKE, BRONSTEIN, ROBERTS
ADDRESSEE: 6 CUSHMAN
STREET: 130 WATER STREET
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: US
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/231,894A
FILING DATE: 22-APR-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/873907
FILING DATE: 24-APR-1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 097655-1991

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/231,894A
FILING DATE: 22-APR-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/873907
FILING DATE: 24-APR-1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 097655-1991
FILING DATE: 26-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 066381-1992
FILING DATE: 24-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: RESNICK, DAVID S.
REGISTRATION NUMBER: 34235

```

1      COMPUTER READABLE FORM:
2      MEDIUM TYPE: Floppy disk
3      COMPUTER: IBM PC compatible
4      OPERATING SYSTEM: PC-DOS/MS-DOS
5      SOFTWARE: Patentin Release #1.0, Version #1.30
6      CURRENT APPLICATION DATA:
7      APPLICATION NUMBER: US/09/105,678A
8      FILING DATE: 26-JUN-1998
9      PRIOR APPLICATION DATA:
10     APPLICATION NUMBER: JP 172118/1997
11     FILING DATE: 27-JUN-1997
12     ATTORNEY/AGENT INFORMATION:
13     NAME: Conlin, David G.
14     REGISTRATION NUMBER: 27,026
15     REFERENCE/DOCKET NUMBER: 48466-342
16     TELECOMMUNICATION INFORMATION:
17     TELEPHONE: 617-523-3400
18     TELEFAX: 617-523-6440
19     INFORMATION FOR SEQ ID NO: 49:
20     SEQUENCE CHARACTERISTICS:
21     LENGTH: 146 amino acids
22     TYPE: amino acid
23     STRANDEDNESS:
24     TOPOLOGY: linear
25     MOLECULE TYPE: peptide
26     OS-09-105-678A-49

```

Query Match 33.18; Score 60; DB 3; Length 146;
 Best Local Similarity 40.08; Pred. No. 0.15;
 Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 3 MCTWDAFYLTVHPQGYIEGVRLHVD 27

DB 24 YCKNGGFELRIHPDGRVDGVREKSD 48

Search completed: November 30, 2001, 09:43:37
 Job time: 196 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model.

Run on: November 30, 2001, 09:42:56 : Search time 70.16 Seconds
(without alignments)
20.060 Million cell updates/sec

Title: US-08-846-234-2
Perfect score: 105
Sequence: 1 PVSVCFCVGFDPASEPDSRH 19

Scoring table: BIOSUM62
Gapop 10.0, Gapext 0.5

Searched: 522463 seqs, 74073290 residues
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
A.Geneseq_1101.*
1: /SID8/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID8/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID8/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID8/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID8/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID8/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID8/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID8/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID8/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID8/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID8/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID8/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID8/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID8/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID8/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID8/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID8/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID8/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID8/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID8/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID8/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID8/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	105	100.0	19	19	AAW53568
2	105	100.0	19	20	AAV17423
3	105	100.0	784	19	AAW53570
4	105	100.0	784	20	AAV17417
5	71	67.6	780	22	AAW57859
6	71	67.6	781	19	AAW57887
7	71	67.6	781	20	AAV30143
8	71	67.6	781	22	AAW49400
9	63	60.0	799	19	AAW57886
10	63	60.0	799	20	AAV30142
11	62	59.0	758	21	AAV70978

12	61	58.1	783	20	AAV32073
13	47.5	45.2	770	21	AAV70976
14	47	44.8	175	21	AAV16865
15	47	44.8	253	21	AAV16864
16	47	44.8	295	21	AAV16863
17	46.5	44.3	777	20	AAV32074
18	45	42.9	763	21	AAV70977
19	44	41.9	112	18	AAW08433
20	43	41.0	313	22	AAW50675
21	43	41.0	441	20	AAV40501
22	43	41.0	447	21	AAV15496
23	42.5	40.5	841	21	AAV70981
24	42	40.0	22	20	AAV13176
25	42	40.0	257	17	AAW94460
26	42	40.0	382	22	AAW90340
27	42	40.0	382	22	AAW76725
28	41	39.0	212	20	AAW82677
29	41	39.0	212	20	AAW82678
30	41	39.0	212	20	AAW82685
31	40	38.1	129	22	AAW00853
32	40	38.1	144	20	AAV59807
33	40	38.1	145	20	AAV48481
34	40	38.1	232	22	AAW82256
35	40	38.1	291	20	AAW78465
36	40	38.1	291	20	AAW78466
37	40	38.1	291	20	AAW78464
38	40	38.1	306	22	AAW81740
39	40	38.1	306	22	AAW82698
40	40	38.1	306	22	AAW82773
41	40	38.1	331	22	AAW39136
42	40	38.1	789	13	AAW23159
43	40	38.1	789	13	AAW23164
44	40	38.1	834	12	AAW3143
45	40	38.1	834	13	AAW23158

ALIGNMENTS

RESULT 1	
ID	AAW53568 standard; peptide: 19 AA.
XX	
AC	AAW53568;
XX	
DT	06-JUL-1998 (first entry)
XX	
DE	Cucumber raffinose synthase residues 61 to 79.
XX	
KW	Cucumber; raffinose synthase; sucrose; galactinol.
XX	
OS	Cucumis sativus.
XX	
PN	JPI0084973-A.
XX	
PD	07-APR-1998.
XX	
PF	28-APR-1997; 97JP-011124.
XX	
PR	26-JUL-1996; 96JP-0198079.
XX	
PR	26-APR-1996; 96JP-0107682.
XX	
PA	(AJIN) AJINOMOTO KK.
XX	
DR	WPI: 1998-264858/24.
XX	
PT	Raffinose synthase gene - useful for preparation of raffinose in transformed plant
XX	
PS	Example 2; Page 17; 26pp: Japanese.
XX	
CC	The present sequence is a cucumber raffinose synthase fragment.
CC	Raffinose synthase forms raffinose from sucrose and galactinol, has

PT New raffinose synthase gene - for production of raffinose from
PT sucrose and galactinol

PS Claim 2; Page 25-27; 37pp; Japanese.

CC The present invention describes a raffinose synthase, having an activity
CC of forming raffinose from sucrose and galactinol. The raffinose synthase
CC gene can be used for expression in a plant for the production of
CC raffinose. The raffinose synthase can give raffinose from sucrose
CC and galactinol efficiently. The present sequence represents raffinose
CC synthase from cucumber.

XX Sequence 784 AA;

Query Match 100.0%; Score 105; DB 20; Length 784;

Best Local Similarity 100.0%; Pred. No. 6.9e-09; Mismatches 0; Indels 0; Gaps 0;

OY 1 PVSVCVFGEFADSEPDsrh 19
Db 61 pvsvcvgyfdaeprsrh 79

RESULT 5

AAB98659 AAB98659 standard; protein; 780 AA.

AC AAB98659;

DT 17-AUG-2001 (first entry)

DE Soybean protein; SEQ ID 1.

XX Mutant; mutant; raffinose synthase; raffinose oligosaccharide reduction;
KM plant; soybean.

XX Glycine max.

PN JP2001078783-A.

PD 27-MAR-2001.

PF 03-JUL-2000; 2000JP-0200571.

PR 09-JUL-1999; 99JP-0196036.

PA (SUMO) SUMITOMO CHEM CO LTD.

DR MPI: 2001-313373/33.

DR N-PSDB; AAH27438.

PT Novel mutant protein of raffinose synthase is useful for reducing the
PT raffinose oligosaccharide content in a plant body -

PS Disclosure; Page 18-20; 30pp; Japanese.

CC The present invention relates to a mutant protein of raffinose synthase
CC in which at least one aromatic amino acid present at the position of
CC about 1-7 amino acids from the N-terminus is deleted or replaced. The
CC mutant protein can be used for reducing the raffinose oligosaccharide
CC content in a plant body. The present protein from soybean, was used in
CC the present invention.

XX Sequence 780 AA;

Query Match 67.6%; Score 71; DB 22; Length 780;

Best Local Similarity 81.2%; Pred. No. 0.0025; Mismatches 3; Indels 0; Gaps 0;

OY 4 VGCVFGEFADSEPDsrh 19
XXXXXXXXXXXXXXXXXXXX

Db 68 vgcfyghadeprsrh 83

RESULT 6

AAW57887 AAW57887 standard; Protein; 781 AA.

AC AAW57887;

DT 23-SEP-1998 (first entry)

DE Soybean raffinose synthetase.

KW Raffinose synthetase; metabolism modification; food additive;

KW gastrointestinal flora; soybean.

OS Glycine max.

PN EP849359-A2.

PD 24-JUN-1998.

PF 18-DEC-1997; 97EP-0122417.

PR 18-DEC-1996; 96JP-0338673.

PA (SUMO) SUMITOMO CHEM CO LTD.

PI Oeda K, Wantanabe E;

DR MPI: 1998-324670/29.

DR N-PSDB; AAV40801.

PT New nucleic acid molecule encoding plant raffinose synthetase -
PT capable of producing raffinose, used as food additives with
PT beneficial effects on gastrointestinal flora

PS Claim 1; Page 31-34; 44pp; English.

CC This sequence represents the soybean raffinose synthetase of the
CC invention. The raffinose synthetase is capable of producing raffinose by
CC combining a D-galactosyl group through an alpha (1-6) bond with a
CC hydroxyl group attached to the carbon atom at position 6 of a D-glucose
CC residue in a sucrose molecule. The DNA can be used to modify metabolism
CC of a host organism by introducing into the host organism or its cell so
CC that the content of the raffinose family oligosaccharides in the host
CC organism or cell is changed. Raffinose oligosaccharides are useful as
CC food additives with beneficial effects on the gastrointestinal flora.

XX Sequence 781 AA;

Query Match 67.6%; Score 71; DB 19; Length 781;

Best Local Similarity 81.2%; Pred. No. 0.0025; Mismatches 3; Indels 0; Gaps 0;

OY 4 VGCVFGEFADSEPDsrh 19
Db 68 vgcfyghadeprsrh 83

RESULT 7

AAV30143 AAV30143 standard; Protein; 781 AA.

AC AAV30143;

DT 26-OCT-1999 (first entry)

DE Amino acid sequence of a raffinose synthase protein.

KW Raffinose synthase; plant; sucrose; raffinose.

XX

```
OS Glycine max.
XX
PN JP11215984-A.
XX
PD 10-AUG-1995.
XX
PF 12-DEC-1997; 97JP-0342899.
XX
PR 28-NOV-1997; 97JP-0329006.
XX
PR 18-DEC-1996; 96JP-0338673.
XX
PA (SUMO ) SUMITOMO CHEM CO LTD.
XX
DR WPI: 1999-511112/43.
XX
DR N-PSDB; AA210002.
XX
PT New raffinose synthase gene - is prepared from a plant material
XX
PS Claim 8; Page 25-27; 40pp; Japanese.
XX
CC The present sequence represents a raffinose synthase protein. The
CC sequence is isolated from plant material. The protein forms raffinose
CC by complexing alpha(1 to 6) - D-galactosyl hydroxyl group of the 6C
CC of D-glucose residue in sucrose molecules.
XX
SQ Sequence 781 AA;

Query Match 67.6%; Score 71; DB 20; Length 781;
Best Local Similarity 81.2%; Pred. No. 0.0025;
Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 4 VGCFVGFDASEPDsrh 19
   ||||| | | | |
Db 68 vgcfvghadeprrsh 83

RESULT 8
AAB49400
ID AAB49400 standard; Protein; 781 AA.
XX
AC AAB49400;
XX
DT 07-MAR-2001 (first entry)
XX
DE Soybean raffinose synthase.
XX
KW Plant promoter; transgenic plant; desired property.
XX
OS Glycine max.
XX
PN EP1048733-A2.
XX
PD 02-NOV-2000.
XX
PF 27-APR-2000; 2000EP-0108962.
XX
PR 30-APR-1999; 99JP-0124527.
XX
PR 01-SEP-1999; 99JP-0247211.
XX
PA (SUMO ) SUMITOMO CHEM CO LTD.
XX
PI Ishige F, Matanabe E, Oeda K;
XX
DR WPI: 2001-104537/12.
XX
DR N-PSDB; AAC89523.
XX
PT New soybean plant promoters useful for generating transgenic plants
XX with desired properties -
XX
PS Example 6; Page 24-27; 36pp; English.
XX
CC The present invention provides novel plant promoters which can be used in
```

```
CC the production of transgenic plants which express genes with desired
CC properties.
XX
SQ Sequence 781 AA;

Query Match 67.6%; Score 71; DB 22; Length 781;
Best Local Similarity 81.2%; Pred. No. 0.0025;
Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 4 VGCFVGFDASEPDsrh 19
   ||||| | | | |
Db 68 vgcfvghadeprrsh 83

RESULT 9
AAW57886
ID AAW57886 standard; Protein; 799 AA.
XX
AC AAW57886;
XX
DT 23-SEP-1998 (first entry)
XX
DE Broad bean raffinose synthetase.
XX
KW Raffinose synthetase; metabolism modification; food additive;
KW gastrointestinal flora; broad bean.
XX
OS Vicia faba.
XX
PN EP849359-A2.
XX
PD 24-JUN-1998.
XX
PF 18-DEC-1997; 97EP-0122417.
XX
PR 18-DEC-1996; 96JP-0338673.
XX
PA (SUMO ) SUMITOMO CHEM CO LTD.
XX
PI Oeda K, Matanabe E;
XX
DR WPI: 1998-324670/29.
XX
DR N-PSDB; AAV40800.
XX
PT New nucleic acid molecule encoding plant raffinose synthetase -
PT capable of producing raffinose, used as food additives with
PT beneficial effects on gastrointestinal flora
XX
PS Claim 1; Page 26-29; 44pp; English.
XX
CC This sequence represents the broad bean raffinose synthetase of the
CC invention. The raffinose synthetase is capable of producing raffinose by
CC combining a D-galactosyl group through an alpha (1-6) bond with a
CC hydroxyl group attached to the carbon atom at position 6 of a D-glucose
CC residue in a sucrose molecule. The DNA can be used to modify metabolism
CC of a host organism by introducing into the host organism or its cell so
CC that the content of the raffinose family oligosaccharides in the host
CC organism or cell is changed. Raffinose oligosaccharides are useful as
CC food additives with beneficial effects on the gastrointestinal flora.
XX
SQ Sequence 799 AA;

Query Match 60.0%; Score 63; DB 19; Length 799;
Best Local Similarity 66.7%; Pred. No. 0.052;
Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 5 GCFVGFDASEPDsrh 19
   ||||| | | | |
Db 87 gcfvghadeprrsh 101
```

```

RESULT 10
AAAY30142
ID AAY30142 standard; Protein; 799 AA.
AC AAY30142;
XX
XX
DT 26-OCT-1999 (first entry)
DE Amino acid sequence of a raffinose synthase protein.
XX
XX
KW Raffinose synthase; plant; broad bean; sucrose; raffinose.
XX
XX
OS Vicia faba.
XX
XX
PN JP1215984-A.
XX
XX
PD 10-AUG-1999.
XX
XX
PF 12-DEC-1997; 97JP-0342899.
XX
XX
PR 28-NOV-1997; 97JP-0329006.
PR 18-DEC-1996; 96JP-0338673.
XX
XX
PA (SUMO ) SUMITOMO CHEM CO LTD.
XX
XX
DR MPI: 1999-511112/43.
DR N-PSDB; AAZ10001.
XX
XX
PT New raffinose synthase gene - is prepared from a plant material
XX
XX
PS Claim 5; Page 19-21; 40pp; Japanese.
XX
XX
CC The present sequence represents a raffinose synthase protein. The
CC sequence is isolated from plant material of broad beans. The
CC protein forms raffinose by complexing alpha(1 to 6) - D-galactosyl
CC hydroxyl group of the 6C of D-glucose residue in sucrose molecules.
XX
XX
SO Sequence 799 AA:

Query Match 60.0%; Score 63; DB 20; Length 799;
Best Local Similarity 66.7%; Pred. No. 0.052;
Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 GCFVGFDPASEPDSRH 19
DB 87 gctvgfnstepksh 101

RESULT 11
AAAY70978
ID AAY70978 standard; Protein; 758 AA.
AC AAY70978;
XX
XX
DT 09-AUG-2000 (first entry)
DE Soybean raffinose synthase from clone sfl1.pk125.d4.
XX
XX
KW Soybean; raffinose synthase; raffinose saccharide;
KW clone sfl1.pk125.d4; nutritional; soy protein.
XX
XX
OS Glycine max.
XX
XX
PN WO200024915-A2.
XX
XX
PD 04-MAY-2000.
XX
XX
PF 22-OCT-1999; 99WO-US24923.
XX
XX
PR 23-OCT-1998; 98US-0105451.
XX
XX
PA (DUPO ) DU PONT DE NEMOURS & CO E I.

```

```

XX
XX
PI Allen SM, Hitz WD;
XX
XX
DR MPI: 2000-350754/30.
DR N-PSDB; AAD00335.
XX
XX
PT Nucleic acids and encoded proteins involved in the biosynthesis of
PT raffinose, useful for producing soybean seeds with a reduced raffinose
PT content and therefore improved nutritional quality -
XX
XX
PS Claim 2; Page 47-49; 58pp; English.
XX
XX
CC The present sequence is a raffinose synthase from
CC clone sfl1.pk125.d4 isolated from a soybean immature flower cDNA
CC library sfl1. Raffinose synthase is involved in the biosynthesis
CC of raffinose and higher homologues in the raffinose saccharide family
CC from sucrose. The present sequence is useful for reducing the raffinose
CC saccharide content of soybean seeds which improves the nutritional
CC quality of the soy protein products derived from them.
XX
XX
SO Sequence 758 AA:

Query Match 59.0%; Score 62; DB 21; Length 758;
Best Local Similarity 58.8%; Pred. No. 0.071;
Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 3 SVCCFVGFDPASEPDSRH 19
DB 54 ttcgflgfnatspksrh 70

RESULT 12
AAAY32073
ID AAY32073 standard; Protein; 783 AA.
AC AAY32073;
XX
XX
DT 17-JAN-2000 (first entry)
XX
XX
DE Sugarbeet raffinose synthase.
XX
XX
KW Raffinose synthase; sugarbeet; transgenic plant.
XX
XX
OS Beta vulgaris.
XX
XX
FH Key Location/Qualifiers
FT Misc-difference 227 /note= "encoded by CCW"
FT Misc-difference 348 /note= "encoded by CCR"
XX
XX
EP953643-A2.
XX
XX
PD 03-NOV-1999.
XX
XX
PF 27-APR-1999; 99EP-0107430.
XX
XX
PR 30-APR-1998; 98JP-0120550.
PR 30-APR-1998; 98JP-0120551.
PR 04-DEC-1998; 98JP-0345590.
PR 10-DEC-1998; 98JP-0351246.
XX
XX
PA (SUMO ) SUMITOMO CHEM CO LTD.
XX
XX
PI Watanabe E, Oeda K;
XX
XX
DR MPI: 1999-593144/51.
DR N-PSDB; AAZ20208.
XX
XX
PT New sense and antisense genes, useful for altering the level of
PT raffinose in food plants -
XX

```


PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.

PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160747.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.

PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
Query Match 44.84; Score 47; DB 21; Length 175;
Best Local Similarity 64.38; Pred. No. 4.5;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 5 GCFVGFDPASEPDSR 18
|:|:|:|:|:|:|
Db 90 geflfgideneptsr 103
RESULT 15
AAG16864
ID AAG16864 standard; Protein; 253 AA.
XX AAG16864;
XX 17-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 17670.
DE
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
OS
PN EPI033405-A2.
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
PE
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142380.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.

PR 06-AUG-1999: 99US-0147416. QY 5 GCPVGFDPASPPDSR 18
PR 09-AUG-1999: 99US-0147493. | :||| :|| |
PR 10-AUG-1999: 99US-0147935. Db 168 getlgtideneptsr 181
PR 11-AUG-1999: 99US-0148319.
PR 12-AUG-1999: 99US-0148341.
PR 13-AUG-1999: 99US-0148565.
PR 13-AUG-1999: 99US-0148684.
PR 16-AUG-1999: 99US-0149368.
PR 17-AUG-1999: 99US-0149375.
PR 18-AUG-1999: 99US-0149426.
PR 20-AUG-1999: 99US-0149722.
PR 20-AUG-1999: 99US-0149723.
PR 20-AUG-1999: 99US-0149929.
PR 23-AUG-1999: 99US-0149902.
PR 23-AUG-1999: 99US-0149930.
PR 25-AUG-1999: 99US-0150566.
PR 26-AUG-1999: 99US-0150884.
PR 27-AUG-1999: 99US-0151065.
PR 27-AUG-1999: 99US-0151066.
PR 27-AUG-1999: 99US-0151080.
PR 30-AUG-1999: 99US-0151303.
PR 31-AUG-1999: 99US-0151438.
PR 01-SEP-1999: 99US-0151930.
PR 07-SEP-1999: 99US-0152363.
PR 10-SEP-1999: 99US-0153070.
PR 13-SEP-1999: 99US-0153758.
PR 15-SEP-1999: 99US-0154018.
PR 16-SEP-1999: 99US-0154039.
PR 20-SEP-1999: 99US-0154779.
PR 22-SEP-1999: 99US-0155139.
PR 23-SEP-1999: 99US-0155486.
PR 24-SEP-1999: 99US-0155659.
PR 28-SEP-1999: 99US-0156458.
PR 29-SEP-1999: 99US-0156596.
PR 04-OCT-1999: 99US-0157117.
PR 05-OCT-1999: 99US-0157753.
PR 06-OCT-1999: 99US-0157865.
PR 07-OCT-1999: 99US-0158029.
PR 08-OCT-1999: 99US-0158232.
PR 12-OCT-1999: 99US-0158369.
PR 13-OCT-1999: 99US-0159293.
PR 13-OCT-1999: 99US-0159294.
PR 13-OCT-1999: 99US-0159295.
PR 14-OCT-1999: 99US-0159329.
PR 14-OCT-1999: 99US-0159330.
PR 14-OCT-1999: 99US-0159331.
PR 14-OCT-1999: 99US-0159637.
PR 14-OCT-1999: 99US-0159638.
PR 18-OCT-1999: 99US-0159584.
PR 21-OCT-1999: 99US-0160741.
PR 21-OCT-1999: 99US-0160767.
PR 21-OCT-1999: 99US-0160768.
PR 21-OCT-1999: 99US-0160770.
PR 21-OCT-1999: 99US-0160814.
PR 21-OCT-1999: 99US-0160815.
PR 22-OCT-1999: 99US-0160980.
PR 22-OCT-1999: 99US-0160981.
PR 22-OCT-1999: 99US-0160989.
PR 25-OCT-1999: 99US-0161404.
PR 25-OCT-1999: 99US-0161405.
PR 25-OCT-1999: 99US-0161406.
PR 26-OCT-1999: 99US-0161359.
PR 26-OCT-1999: 99US-0161360.
PR 26-OCT-1999: 99US-0161361.
PR 28-OCT-1999: 99US-0161920.
PR 28-OCT-1999: 99US-0161992.
PR 28-OCT-1999: 99US-0161993.
PR 29-OCT-1999: 99US-0162142.

Search completed: November 30, 2001, 09:42:57
Job time: 156 sec

Query Match 44.8%; Score 47; DB 21; Length 253;
Best Local Similarity 64.3%; Pred. No. 6.5;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 30, 2001, 09:41:40 ; Search time 39.46 Seconds
(without alignments)
36.678 Million cell updates/sec

Title: US-08-846-234-2

Perfect score: 105

Sequence: 1 PVSVCVFVGFDPSPSRH 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR68:*
2: pir1:*
3: pir2:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	47	44.8	295	2	B66320
2	47	44.8	295	2	T51362
3	44	41.9	112	2	S68584
4	44	41.9	301	2	S73023
5	44	41.9	428	2	A43741
6	44	41.9	498	2	T48269
7	44	41.9	650	1	UC1450
8	43	41.0	313	2	T33185
9	43	41.0	441	2	F86185
10	43	41.0	447	2	T02547
11	43	41.0	464	2	S59898
12	43	41.0	464	2	T48675
13	43	41.0	478	2	T15516
14	42.5	40.5	546	2	A6890
15	42	40.0	72	2	T46623
16	42	40.0	169	2	G72559
17	42	40.0	240	2	E83896
18	42	40.0	311	2	T32776
19	42	40.0	314	2	T27383
20	42	40.0	511	2	T23459
21	41.5	39.5	219	2	T42605
22	41	39.0	125	2	E81192
23	41	39.0	314	2	T25842
24	41	39.0	552	2	T16345
25	41	39.0	533	2	A45086
26	40.5	38.6	498	2	A83635
27	40	38.1	80	2	G86417
28	40	38.1	104	2	S36510
29	40	38.1	260	2	G86228

30	40	38.1	282	2	A36875	plasmalemma vola
31	40	38.1	462	2	A84187	hypothetical prote
32	40	38.1	447	2	T50705	gamma-aminobutyrat
33	40	38.1	465	2	G02652	kyureninase (EC 3
34	40	38.1	609	2	E82423	hypothetical prote
35	40	38.1	615	1	ABCHS	serum albumin prec
36	40	38.1	650	2	A54976	peroxisomal matrix
37	40	38.1	755	2	B41836	amine oxidase (fla
38	40	38.1	761	1	IRBODE	desmocollin Ia - b
39	40	38.1	799	2	S18209	fibroblast growth
40	40	38.1	839	1	IBBODF	desmocollin Ib pre
41	39.5	37.6	329	2	F64356	translation initia
42	39.5	37.6	405	1	VCBEGF	glycoprotein G pre
43	39.5	37.6	435	2	T42613	probable envelope
44	39.5	37.6	1208	2	B82091	exodeoxyribonuclea
45	39	37.1	70	2	I40186	hypothetical prote

ALIGNMENTS

RESULT 1
B66320
3-phosphoserine phosphatase [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: B66320
R:JTheologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon, Chiu, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, Ansen, N.F.; Hughes, B.; Hulsar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Malt, R.; Maritz Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Talloker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: B66320
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-295 <STO>
A:Cross-references: GB:AE005172; NID:g9795592; PIDN:AAF98410.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match 44.8%; Score 47; DB 2; Length 295;
Best Local Similarity 64.3%; Pred. No. 3.7;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 GCFVGFDPSPSR 18
DB 210 GEFVGFDPSPSR 223

RESULT 2
T51362
3-phosphoserine phosphatase (EC 3.1.3.3) precursor, chloroplast [validated] - Arabidops
N:Alternate names: 3-phosphoserine phosphatase
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 18-Aug-2000
C:Accession: T51362
R:Ho, C.; Noji, M.; Saito, K.
J. Biol. Chem. 274, 11007-11012, 1999
A:Title: Plastidic pathway of serine biosynthesis. Molecular cloning and expression o
A:Reference number: 225385
A:Accession: T51362
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-295 <HOC>
A:Cross-references: EMBL:AB018408; PIDN:BA03806.1
C:Genetics:

C:Species: Rattus norvegicus (Norway rat)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 01-Dec-2000
C:Accession: J01450; PT0191
R:Horlick, R.A.; Stock, S.L.; Cooke, G.M.
Gene 120, 291-295, 1992
A:Title: Cloning, expression and tissue distribution of the gene encoding rat fibroblast
A:Reference number: J01450; MUID:93013049
A:Accession: J01450
A:Molecule type: mRNA
A:Residues: 1-650 <HOR>
A:Cross-references: GB:M91599; NID:g204137; PIDN:AAA1157.1; PID:g204138
R:Li, C.; Lemke, G.
Neuron 6, 691-704, 1991
A:Title: An extended family of protein-tyrosine kinase genes differentially expressed in
A:Reference number: PT0191
A:Accession: PT0191
A:Molecule type: mRNA
A:Residues: 465-518 <LA1>
A:Experimental source: sciatic nerve
C:Genetics:
A:Gene: FGFR4; tyro-9
C:Function:
A:Description: receptor mediating effects of fibroblast growth factor
A>Note: expressed in normal lung; expressed in some carcinomas
C:Superfamily: basic fibroblast growth factor receptor 1; immunoglobulin homology; prot
C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; growth factor receptor;
F:11-72/Domain: immunoglobulin homology <IM1>
F:110-161/Domain: immunoglobulin homology <IM2>
F:218-238/Domain: transmembrane #status predicted <TM>
F:239-650/Domain: intracellular #status predicted <INT>
F:313-329/Domain: protein kinase homology <KIN>
F:321-329/Region: protein kinase ATP-binding motif
F:104_136_157_168/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:351_368_460/Active site: Lys, Glu, Asp #status predicted
F:465_478/Binding site: magnesium (Asn, Asp) #status predicted
F:491/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #status predicte

Query Match 41.9%; Score 44; DB 1; Length 650;
Best Local Similarity 37.5%; Pred. No. 26;
Matches 9; Conservative 3; Mismatches 4; Indels 8; Gaps 1;
OY 1 PVSVCF-----VGFDPSPD 16
1: ||||| : |||||
DB 320 PEGECFGVCAEALGMDSSRPD 343

RESULT 8
T33185
hypothetical protein F22F7.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T33185
R:Miller, N.; Kramer, J.; Smith, A.
submitted to the EMBL Data Library, May 1998
A:Description: The sequence of C. elegans cosmid F22F7.
A:Reference number: Z21299
A:Residues: T33185
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-313 <MIL>
A:Cross-references: EMBL:AF067937; PIDN:AAC19217.1; GSPDB:GN00023; CESP:F22F7.6
A:Experimental source: strain Bristol N2; clone F22F7
C:Genetics:
A:Gene: CESP:F22F7.6
A:Map position: 5
A:introns: 18/2; 82/3; 117/1; 210/2; 241/3; 287/2

Query Match 41.0%; Score 43; DB 2; Length 313;
Best Local Similarity 50.0%; Pred. No. 18;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 1 PVSVCFGVDPSPD 14
1: ||||| : |||||
DB 51 PISVGFIDFHSSQ 64

RESULT 9
F86185
hypothetical protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: F86185
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Malt, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: F86185
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-441 <STO>
A:Cross-references: GB:AE005172; NID:g2388583; PIDN:AA871464.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match 41.0%; Score 43; DB 2; Length 441;
Best Local Similarity 57.1%; Pred. No. 26;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
OY 4 VCGVGFDPSPD 17
1: ||||| : |||||
DB 139 LGFVGFPDNDPDS 152

RESULT 10
T02547
hypothetical protein At2g32480 [imported] - Arabidopsis thaliana
N:Alternate names: hypothetical protein T26B15.4
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 16-Feb-2001
C:Accession: T02547; F84733
R:Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Syke
submitted to the EMBL Data Library, July 1998
A:Description: Arabidopsis thaliana chromosome II BMC T26B15 genomic sequence.
A:Reference number: Z14678
A:Accession: T02547
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-447 <ROU>
A:Cross-references: EMBL:AC004681; NID:g3298532; PID:g3298536
A:Experimental source: cultivar Columbia
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
M.; Koo, H.; Moffat, K.S.; Conlin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, T.; Tallon
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: F84733
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-447 <STO>
A:Cross-references: GB:AE002093; NID:g3298536; PIDN:AAC25930.1; GSPDB:GN00139
C:Genetics:
A:Gene: T26B15.4; At2g32480
A:Map position: 2
C:Superfamily: Escherichia coli probable zinc proteinase yael

C:Accession: T46623
R:Chang, S.; Purvey, J.; Funkhouser, E.A.; Newton, R.J.; Cairney, J.
submitted to the EMBL Data Library, July 1995
A:Description: Cloning of a chitinase homolog which lacks chitin binding sites and is do
A:Reference number: 223105
A:Accession: T46623
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-72 <CHA>
A:Cross-references: EMBL:U31309; NID:9974285; PID:9974288
A:Experimental source: strain s6PT2xs6PT3; 8 month seedlings

Query Match 40.0%; Score 42; DB 2; Length 72;
Best Local Similarity 63.6%; Pred. No. 5.9;
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 6 CPVGFDPASEPD 16
|||
DB 16 CPAGFCGKEPD 26

Search completed: November 30, 2001, 09:41:42
Job time: 81 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 30, 2001, 09:40:54 ; Search time 23.6 Seconds

(without alignments)
29.518 Million cell updates/sec

Title: US-08-846-234-2

Perfect score: 105

Sequence: 1 PVSVCFCVGFDPSEPSRH 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 3664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : SWISSProt_39:*

SUMMARIES

Result No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	44	41.9	112	1	HYPL_AGABI
2	44	41.9	428	1	TERM_DROME
3	43	41.0	464	1	KYNU_RAT
4	43	41.0	478	1	KYNU_CAEL
5	41.5	39.5	222	1	GLI9_ARATH
6	41	39.0	5217	1	HTSL_COCCA
7	40	38.1	104	1	VE7_HPV32
8	40	38.1	282	1	PORL_BOVIN
9	40	38.1	282	1	PORL_RABIT
10	40	38.1	296	1	PORL_MOUSE
11	40	38.1	367	1	PKK_PARRP
12	40	38.1	465	1	KYNU_HUMAN
13	40	38.1	615	1	ALBU_CHICK
14	40	38.1	650	1	PEX8_PICAN
15	40	38.1	755	1	AMO_RJEA
16	40	38.1	834	1	DPOL_THECA
17	40	38.1	834	1	DPOL_THETH
18	40	38.1	893	1	DSCL_BOVIN
19	39.5	37.6	329	1	E2B1_METJA
20	39.5	37.6	405	1	VGL4_HSV4
21	39.5	37.6	415	1	RL3_DROME
22	39	37.1	301	1	MCAT_HUMAN
23	39	37.1	331	1	DEFA_MOUSE
24	39	37.1	389	1	BMR1_BACSU
25	39	37.1	481	1	KRL_HSV11
26	39	37.1	500	1	YK23_CAEL
27	39	37.1	725	1	YGL4_YEAST
28	39	37.1	752	1	BS11_TRYCR
29	39	37.1	757	1	AMO_ECOLI
30	39	37.1	1088	1	RPO_ROTBR
31	39	37.1	1088	1	RPO_ROTBU
32	39	37.1	1088	1	RPO_ROTGP
33	39	37.1	1088	1	RPO_ROTSL

34	39	37.1	1337	1	P152_YEAST
35	38.5	36.7	218	1	VGL4_HSVB
36	38.5	36.7	506	1	CP4B_RABIT
37	38.5	36.7	1074	1	SM5A_HUMAN
38	38.5	36.7	1077	1	SM5A_MOUSE
39	38	36.2	111	1	VE7_HPV07
40	38	36.2	216	1	CSGD_SALTY
41	38	36.2	301	1	MCAT_RAT
42	38	36.2	475	1	TRB1_ECOLI
43	38	36.2	481	1	KRL_HSV2H
44	38	36.2	495	1	PORL_SYNT3
45	38	36.2	702	1	SYT_MYCLE

ALIGNMENTS

RESULT 1
HYPL_AGABI STANDARD: PRT: 112 AA.
AC P49072: 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYDROPHOBIN 1 PRECURSOR (HYDROPHOBIN A).
GN HYPA OR ABH1.
OS Agaricus bisporus (Common mushroom).
OC Agaricota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Agaricales; Agaricaceae; Agaricus.
OX NCBI_TaxID=5341;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HORST H39;
RA MEDLINE=96192085; PubMed=8632464;
RX de Groot P.W.J., Schaap P.J., Visser J., van Griensven L.J.L.D.;
RT "The Agaricus bisporus hypha gene encodes a hydrophobin and
RT specifically accumulates in peel tissue of mushroom caps during fruit
RT body development";
RL J. Mol. Biol. 257:1008-1018(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HORST U3;
RX MEDLINE=96254018; PubMed=8704971;
RA Lugones L.G., Bosscher J.S., Scholtmeyer K., de Vries O.M.H.,
RA Wessels J.G.H.;
RT "An abundant hydrophobin (ABH1) forms hydrophobic rodlet layers in
RT Agaricus bisporus fruiting bodies.";
RL Microbiology 142:1321-1329(1996).
CC -1- FUNCTION: FORMS HYDROPHOBIC RODLET LAYERS.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN FRUITING BODIES.
CC -1- SIMILARITY: BELONGS TO THE FUNGAL HYDROPHOBIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X89242; CAA61530.1; -;
DR EMBL: X90818; CAA62331.1; -;
DR EMBL: X92861; CAA63447.1; -;
DR InterPro: IPR001338; Hydrophobin.
DR Pfam: PF01185; Hydrophobin; 1.
DR SMART: SM00075; HYDRO: 1.
DR PROSITE: PS00956; HYDROPHOBIN: 1.
KW Fruiting body; Cell wall; Glycoprotein; Signal.
FT SIGNAL 1 19
FT CHAIN 20 112 HYDROPHOBIN 1.
FT CONFLICT 40 41 CD -> Y (IN REF. 2).
FT SEQUENCE 112 AA; 11190 MW; F7647C12826FF6E1 CRC64;

[illegible]

```

RA  Gibbs R A, Myers E.W., Rubin G.M., Venter J.C.;
RT  "The genome sequence of Drosophila melanogaster.";
RL  Science 287;2185-2195(2000).
CC  -1- DEVELOPMENTAL STAGE: EXPRESSED FROM THE CELLULAR BIASTOERM STAGE
CC  ON, MOST DURING GASTROULATION AND IS NO LONGER DETECTED BY THE END
CC  OF GERM BAND EXTENSION.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL: M19140; AAA28928.1; -.
DR  EMBL: AE003520; AAP49257.1; -.
DR  PIR: A43741; A43741.
DR  FLYbase; FBgn0003683; term
DR  InterPro; IPR000822; Znf-C2H2.
DR  PROSITE; PS50157; ZINC_FINGER_C2H2_2, UNKNOWN_1.
DR  DNN-binding; zinc-finger; Developmental protein.
FT  ZN.FING 325 346 C3H-type.
FT  FT 144 144 A -> V (IN REF. 1).
SQ  SEQUENCE 428 AA; 49165 MW; F5456533F738AE434 CRC64;

```

Query Match	41.9%	Score 44	DB 1	Length 428
Best Similarity	50.0%	Pred. No. 7.7		
Best Local	0%			
Matches	7	Conservative	2	Mismatches 5; Indels 0; Gaps 0.
QY	6	CFVGFDASEPDSRH	19	
	1	:	1	:
	1	:	1	:
db	37	CYSIDADEPDSQH	50	

RESULT	3			
ID	KYNU_RAT	STANDARD;	PRT;	464 AA.
AC	KYNU_RAT	P07012; 09QOW90;		
DC		15-DEC-1998 (Rel. 37, Created)		
DT		15-DEC-1998 (Rel. 37, Last sequence update)		
DT		20-AUG-2001 (Rel. 40, Last annotation update)		
DE		KYNURENINASE (EC 3.7.1.3) (L-KYNURENINE HYDROLASE).		
GN		Kynu.		
OS		Rattus norvegicus (Rat).		
OC		Eumaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
OX		NCBI_TaxID=10116;		
RP		[1]		
RP		SEQUENCE, AND MASS SPECTROMETRY.		
RC		STRAIN=H1STAR; TISSUE=Liver;		
RX		MEDLINE=96049498; Pubmed=7578221;		
RA		Takeuchi F., Tsubouchi R., Yoshino M., Shibata Y.;		
RT		"Amino-acid sequence of rat liver kynureninase.";		
RL		Biochim. Biophys. Acta 1252:185-186(1995).		
RN		[2]		
RP		SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.		
RC		TISSUE=Liver;		
RX		MEDLINE=97324088; Pubmed=9180257;		
RA		Toma S., Nakamura M., Tone S., Okuno E., Kido R., Breton J.;		
RA		Avanzi N., Cozzi L., Speciale C., Mostardini M., Gatti S., Benatti L.;		
RT		"Cloning and recombinant expression of rat and human kynureninase.";		
RL		FEBS Lett. 408:5-10(1997).		
RN		[3]		
RP		SEQUENCE OF 19-117 FROM N.A., AND PARTIAL SEQUENCE.		
RC		TISSUE=Liver, and Kidney;		
RX		MEDLINE=96314506; Pubmed=7806755;		
RA		Alberati-Giani D., Buchli R., Malherbe P., Broger C., Lang G.;		
RA		Koehler C., Lahm H.-W., Cesura A.M.;		
RT		"Isolation and expression of a cDNA clone encoding human		
RT		kynureninase.";		
RL		Eur. J. Biochem. 239:460-468(1996).		

```

CC -1- FUNCTION: CATALYZES THE CLEAVAGE OF L-KYURENINE (L-KYN) AND L-3-
CC HYDROXYANTHRANILIC ACIDS (3-OH-KYN), RESPECTIVELY. HAS A PREFERENCE
CC FOR THE L-3-HYDROXY FORM. OPTIMUM ACTIVITY IS AROUND PH 9.0 FOR L-
CC KYN AND AROUND 8.5 FOR L-3OH-KYN. ALSO HAS CYSTEINE-CONJUGATE-
CC BETA-LYASE ACTIVITY.
CC -1- CATALYTIC ACTIVITY: L-KYURENINE + H(2)O = ANTHRANILATE +
CC L-ALANINE.
CC -1- CATALYTIC ACTIVITY: L-3-HYDROXYKYURENINE + H(2)O = 3-
CC HYDROXYANTHRANILATE + L-ALANINE.
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.
CC -1- ENZYME REGULATION: INHIBITED BY O-METHYLBENZOTYALANINE (OMBA).
CC -1- PATHWAY: INVOLVED IN THE BIOSYNTHESIS OF NAD COFACTORS FROM
CC TRYPTOPHAN THROUGH THE KYURENINE PATHWAY.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- TISSUE SPECIFICITY: HIGH LEVELS IN LIVER AND KIDNEY. ALSO
CC DETECTED IN HEART, RETINA, OVARY, LUNG, TESTIS AND BRAIN.
CC -1- INDUCTION: INHIBITED BY THIOL REAGENTS AND HEAVY METAL IONS.
CC -1- SIMILARITY: BELONGS TO THE KYURENINASE FAMILY. SLIGHTLY RELATED
CC TO CLASS-V OF PYRIDOXAL-PHOSPHATE-DEPENDENT AMINOTRANSFERASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U68168; AAC53206.1; -.
CC KW Hydrolyase; Pyridoxal phosphate; Acetylation.
CC FT MOD_RES 1 1 ACETYLATION.
CC FT BINDING 276 276 PYRIDOXAL PHOSPHATE.
CC FT CONFLICT 18 18 A -> T (IN REF. 1).
CC FT CONFLICT 26 26 D -> N (IN REF. 1).
CC FT CONFLICT 118 118 T -> S (IN REF. 1).
CC FT SEQUENCE 464 AA; 52453 MW; FFLCC95E3202EECC CRC64;
CC
CC Query Match 41.0%; Score 43; DB 1; Length 464;
CC Best Local Similarity 100.0%; Pred. No. 12;
CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
CC QY 5 GCFVGF 11
CC DB 244 GCFVGF 250
CC
CC RESULT 4
CC KINU_CAEEL STANDARD; PRT; 478 AA.
CC ID KINU_CAEEL Q18026;
CC DT 15-DEC-1998 (Rel. 37, Created)
CC DT 15-DEC-1998 (Rel. 37, Last sequence update)
CC DT 15-DEC-1998 (Rel. 37, Last annotation update)
CC DE PROBABLE KNURENINASE (EC 3.7.1.3) (L-KYURENINE HYDROLASE).
CC GN C15H9.7.
CC OS Caenorhabditis elegans.
CC OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
CC OC Rhabditidae; Peloderinae; Caenorhabditis.
CC OK NCBI_TaxID=6239;
CC RN
CC RP SEQUENCE FROM N.A.
CC RC STRAIN-BRISTOL N2;
CC RA Bentley D.;
CC RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: L-KYURENINE + H(2)O = ANTHRANILATE +
CC L-ALANINE.
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE KYURENINASE FAMILY. SLIGHTLY RELATED
CC TO CLASS-V OF PYRIDOXAL-PHOSPHATE-DEPENDENT AMINOTRANSFERASES.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U56965; AAB52668.1; -.
CC DR Wormpep; C15H9.7; CE06835.
CC KW Hypothetical protein; Hydrolyase; Pyridoxal phosphate.
CC FT BINDING 289 289 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC FT SEQUENCE 478 AA; 54049 MW; E889450929CE94BD CRC64;
CC
CC Query Match 41.0%; Score 43; DB 1; Length 478;
CC Best Local Similarity 100.0%; Pred. No. 13;
CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
CC QY 5 GCFVGF 11
CC DB 257 GCFVGF 263
CC
CC RESULT 5
CC GL19_ARATH STANDARD; PRT; 222 AA.
CC ID GL19_ARATH Q9FMB0;
CC AC Q9FMB0;
CC DT 20-AUG-2001 (Rel. 40, Created)
CC DT 20-AUG-2001 (Rel. 40, Last sequence update)
CC DT 20-AUG-2001 (Rel. 40, Last annotation update)
CC DE PUTATIVE GERMIN-LIKE PROTEIN SUBFAMILY 1 MEMBER 9 PRECURSOR.
CC GN A7G538910 OR K156.14 OR K156.90.
CC OS Arabidopsis thaliana (Mouse-ear cress).
CC OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
CC OC Spermatophyta; Magnoliophyta; Eudicotyledons; core eudicots; Rosidae;
CC OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
CC OK NCBI_TaxID=3702;
CC RN
CC RP SEQUENCE FROM N.A.
CC RC STRAIN-cv. Columbia;
CC RX MEDLINE-21016721; PubMed-11130714;
CC RA Tabata S., Kaneko T., Nakamura Y., Kotani H., Kato T., Asamizu E.,
CC RA Miyajima N., Sasamoto S., Kimura T., Hosouchi T., Kawashima K.,
CC RA Kohara M., Matsumoto M., Matsuno A., Muraki A., Nakayama S.,
CC RA Nakazaki N., Naruo K., Okumura S., Shimpo S., Takeuchi C., Wada T.,
CC RA Watanabe A., Yamada M., Yasuda M., Sato S., de la Bastide M.,
CC RA Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston R.,
CC RA Habermann K., Murray J., Johnson D., Kohlfing T., Nelson J.,
CC RA Stoneking T., Pepin K., Spieth J., Sekhon M., Armstrong J., Becker M.,
CC RA Belter E., Cordum H., Cordes M., Courtney L., Courtney W., Dante M.,
CC RA Du H., Edwards J., Fryman J., Haakensen B., Lamar E., Latreille P.,
CC RA Leonard S., Meyer R., Mulvaney E., Ozersky P., Riley A., Strommatt C.,
CC RA Wagner-McPherson C., Wollam A., Yeakum M., Bell M., Dedhia N.,
CC RA Parnell L., Shah R., Rodriguez M., Hoon See L., Vil D., Baker J.,
CC RA Kirchhoff K., Toth K., King L., Bahret A., Miller B., Maria M.,
CC RA Martensen R., McCombie W.R., Wilson R.K., Murphy G., Bancroft I.,
CC RA Volckaert G., Wambutt R., Duesterhoeft A., Stiekema W., Pohl T.,
CC RA Ertlan K.-D., Terryn N., Hartley N., Bent E., Johnson S.,
CC RA Langham S.-A., McCullagh B., Robben J., Grynoprez B., Zimmermann W.,
CC RA Ranspberger U., Wedler H., Balke K., Wedler E., Peters S.,
CC RA van Staveren M., Dirks W., Moolman P., Klein Lankhorst R.,
CC RA Welzenegger T., Bothé G., Rose M., Hauf J., Berner S., Hempel S.,
CC RA Feldpausch M., Lamberth S., Villarroel R., Gietlen J., Ardiles W.,
CC RA Bents O., Lemcke K., Kolesov G., Mayer K., Rudd S., Schoof H.,
CC RA Schueller C., Zaccaria P., Mewes H.-W., Beyan M., Franz P.F.;
CC RT "Sequence and analysis of chromosome 5 of the plant Arabidopsis
CC thaliana."
CC RL Nature 408:823-826(2000).
CC RN
CC RP SEQUENCE FROM N.A.
CC RC STRAIN-cv. Columbia;

```

RA MEDLINE-98290546; PubMed-9628562;
RA Sato S., Kaneko T., Kotani H., Nakamura Y., Asamizu E., Miyajima N.,
RA Tabata S.;
RT *Structural analysis of Arabidopsis thaliana chromosome 5. IV.
RT Sequence features of the regions of 1,456,315 bp covered by nineteen
RT physically assigned P1 and TAC clones.";
RL DNA Res. 5:41-54(1998).
CC -1- FUNCTION: MAY PLAY A ROLE IN PLANT DEFENSE. HAS PROBABLY NO
CC OXALATE OXIDASE ACTIVITY EVEN IF THE ACTIVE SITE IS CONSERVED.
CC -1- SUBUNIT: OLIGOMER (BELIEVED TO BE A PENTAMER BUT PROBABLY
CC HEXAMER) (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: APOPLAST (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE GERMIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB009048; BAB08648.1; -
DR InterPro: IPR001929; Germin.
DR Pfam: PF01072; Germin. 1.
DR PRINTS: PR00325; GERMIN. 1.
DR PROSITE: PS00725; GERMIN. 1.
KM Apoplast; Cell wall; Signal; Glycoprotein; Manganese; Metal-binding;
KM Multigene family; Hypothetical protein.
FT SIGNAL 1 222
FT CHAIN 1 222
FT METAL 111 111
FT METAL 113 113 MANGANESE (BY SIMILARITY).
FT METAL 118 118 MANGANESE (BY SIMILARITY).
FT METAL 159 159 MANGANESE (BY SIMILARITY).
FT DISULFID 32 49
FT CARBOHYD 78 78 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 222 AA; 23884 MW; BB925C41A905A4E CRC64;

QY 2 VSVG-CFVGFDPSEPSR 18
1:11 1111 1 1:1
Db 122 VAVGTLPFGFVTSNPNR 139

Query Match 39.5%; Score 41.5; DB 1; Length 222;
Best Local Similarity 55.6%; Pred. No. 10;
Matches 10; Conservative 3; Mismatches 4; Indels 1; Gaps 1;

ID HTSL COCCA STANDARD; PRT; 5217 AA.
AC 001886;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HC-TOXIN SYNTHETASE (EC 6.3.2.-) (HTS).
GN HTSL.
OS Cochliobolus carbonum (Bipolaris zeicola).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes;
OC Pleosporales; Pleosporaceae; Cochliobolus.
OX NCBI_TaxID=5017;
[1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=SB111;
RA MEDLINE-93100328; PubMed-1281482;
RA Scott-Craig J.S., Panackal D.G., Pecard J.-A., Walton J.D.;
RT "The cyclic peptide synthetase catalyzing HC-toxin production in the
RT filamentous fungus Cochliobolus carbonum is encoded by a
RT 15.7-kilobase open reading frame."
RL J. Biol. Chem. 267:26044-26049(1992).
CC -1- FUNCTION: THIS PROTEIN IS A MULTIFUNCTIONAL ENZYME, ABLE TO

CC ACTIVATE THE AMINO ACIDS PRO, L-ALA, D-ALA AND AEO (2-AMINO-9,10-
CC EPOXI-8-OXODECANOIC ACID), AND EPIMERIZE PRO AND L-ALA. IT
CC CATALYZES THE PRODUCTION OF HC-TOXIN: A CYCLIC TETRAPEPTIDE.
CC HTS CONSISTS OF HTS-1 AND HTS-2 WHICH ARE PART OF THE SAME
CC POLYPEPTIDE. HTS-1 ACTIVATES L-PROLINE, THIOESTERIFIES L-PRO, AND
CC EPIMERIZES IT TO D-PRO; HTS-2 ACTIVATES AND THIOESTERIFIES L-ALA
CC AND D-ALA AND EPIMERIZES L-ALA TO D-ALA.
CC -1- COFACTOR: CONTAINS A COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC -1- PATHWAY: NON-RIBOSOMAL BIOSYNTHESIS OF HC-TOXIN.
CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC FAMILY.
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-3 IS THE INITIATOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M96024; AAA33023.1; -
DR HSSP: P14687; 1AMU.
DR InterPro: IPR000873; AMP-bind.
DR InterPro: IPR001242; DUF4.
DR InterPro: IPR003880; Phosphopant_attach.
DR Pfam: PR00668; Condensation; 5.
DR Pfam: PF00550; pp-binding; 4.
DR PRINTS: PR00154; AMPBINDING.
DR PROSITE: PS00012; PHOSPHOPANTETHEINE; 4.
DR PROSITE: PS00455; AMP-BINDING; 3.
DR PROSITE: PS50075; ACP DOMAIN; 4.
KM Ligase; Multifunctional enzyme; Phosphopantetheine; Repeat.
FT REPEAT 249 842
FT REPEAT 1854 2452 DOMAIN 1.
FT REPEAT 3006 3606 DOMAIN 2.
FT REPEAT 4158 4738 DOMAIN 3.
FT DOMAIN 762 840 DOMAIN 4.
FT DOMAIN 2384 2450 ACYL CARRIER (ACP) 1.
FT DOMAIN 3536 3604 ACYL CARRIER (ACP) 2.
FT DOMAIN 4667 4736 ACYL CARRIER (ACP) 3.
FT DOMAIN 803 803 ACYL CARRIER (ACP) 4.
FT BINDING 2414 2414 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 2414 2414 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 3568 3568 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 4700 4700 PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ SEQUENCE 5217 AA; 574637 MW; 0331D9C5400163A5 CRC64;

QY 5 GCFVGFDPSEPSR 18
1111 1 1 1
Db 4200 GCFVLPDSYPHER 4213

Query Match 39.0%; Score 41; DB 1; Length 5217;
Best Local Similarity 57.1%; Pred. No. 3.1e+02;
Matches 8; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

ID VE7 HPV32 STANDARD; PRT; 104 AA.
AC P36827;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE E7 PROTEIN.
GN E7.
OS Human papillomavirus type 32.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=10612;
[1]
RP SEQUENCE FROM N.A.

RA MEDLINE-94265501; PubMed-8205838;

RA "Primer-directed sequencing of human papillomavirus types.";
RA "Cur. Top. Microbiol. Immunol. 186:13-31(1994)."
CC CUR. TOP. MICROBIOL. IMMUNOL. 186:13-31(1994).

CC -1- FUNCTION: E7 PROTEIN HAS BOTH TRANSFORMING AND TRANS-ACTIVATING
CC ACTIVITIES.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: X74475; CAA52550.1; -.
CC PIR: S36510; S36510.
CC InterPro: IPR000148; Papv1_E7.

DR Pfam: PR00527; E7.1.
KW Early protein, transcription regulation; Transforming protein;

FT DNA-binding; Trans-acting factor.
FT DOMAIN 64 67 C-X-X-C MOTIF 1.
FT DOMAIN 97 100 C-X-X-C MOTIF 2.

SO SEQUENCE 104 AA; 11591 MW; F9DFC0B6D804FF CRC64;

Query Match 38.1%; Score 40; DB 1; Length 104;
Best Local Similarity 43.8%; Pred. No. 8.3;
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 1 PVSVCFCVGFDPASEPD 16
Db 28 PVDLYCEQGFQDSD 43

RESULT 8
PORT_BOVIN STANDARD; PRT: 282 AA.

AC P45879;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE VOLTAGE-DEPENDENT ANION-SELECTIVE CHANNEL PROTEIN 1 (VDAC-1)
DE (PLASMALEMMAL PORIN) (BRAIN-DERIVED VOLTAGE-DEPENDENT ANION CHANNEL 1)
DE (BRI-VDAC).
GN VDAC1.

OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.

NCBI_TaxID=9913;
[1]
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RC TISSUE=Brain;
RX MEDLINE-94119914; PubMed-7507248;
RA Dermietzel R., Huang T.-K., Buettner R., Hofer A., Dotzler E.,
RA Krenner M., Deutzmann R., Thümes F.P., Fishman G.I., Spray D.C.,
RA Siemen D.;

RT "Cloning and in situ localization of a brain-derived porin that
RT constitutes a large-conductance anion channel in astrocytic plasma
RT membranes".
RT Proc. Natl. Acad. Sci. U.S.A. 91:499-503(1994).

CC -1- FUNCTION: FORMS A CHANNEL THROUGH THE MITOCHONDRIAL OUTER MEMBRANE
CC AND ALSO THE PLASMA MEMBRANE. THE CHANNEL ALLOWS DIFFUSION OF
CC SMALL HYDROPHILIC MOLECULES; IT ADOPTS AN OPEN CONFORMATION AT LOW
CC OR ZERO MEMBRANE POTENTIAL AND A CLOSED CONFORMATION AT POTENTIALS
CC ABOVE 30-40 MV. THE OPEN STATE HAS A WEAK ANION SELECTIVITY
CC WHEREAS THE CLOSED STATE IS CATION-SELECTIVE.

CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE OF MITOCHONDRIA AND PLASMA
CC MEMBRANE.

CC -1- TISSUE SPECIFICITY: PREDOMINANTLY IN BRAIN ASTROCYTES.
CC -1- DOMAIN: CONSISTS MAINLY OF MEMBRANE-SPANNING SIDED BETA-SHEETS.
CC -1- SIMILARITY: BELONGS TO THE EUKARYOTIC MITOCHONDRIAL PORIN FAMILY.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: X75068; CAA52962.1; -.
CC InterPro: IPR001925; Euk_porin.

DR Pfam: PF01459; Euk_porin.1.
DR PRINTS: PR00185; EUKARYTPORIN.

DR PROSITE: PS00558; EUKARYOTIC_PORIN; 1.
KW Plasma; Porin; Acetylation.

FT INIT_MET 0 0 ACETYLATION (BY SIMILARITY).
FT MOD_RES 1 1
FT MOD_RES 1 1
SO SEQUENCE 282 AA; 30694 MW; 2E6B8205FCD7849 CRC64;

Query Match 38.1%; Score 40; DB 1; Length 282;
Best Local Similarity 41.2%; Pred. No. 23;
Matches 7; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY 2 VSVGCFVGFDPASEPDSR 18
Db 122 INIGCDVDFDIAGPSIR 138

RESULT 9
PORT_RABBIT STANDARD; PRT: 282 AA.

AC O97T15;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE VOLTAGE-DEPENDENT ANION-SELECTIVE CHANNEL PROTEIN 1 (VDAC-1) (OUTER
DE MITOCHONDRIAL MEMBRANE PROTEIN 1).
GN VDAC1.

OS Eucytolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.

NCBI_TaxID=9986;
[1]
SEQUENCE FROM N.A.

RC STRAIN=NEW ZEALAND WHITE; TISSUE=Corneal endothelium;
RA Rae J.L.;

CC -1- FUNCTION: FORMS A CHANNEL THROUGH THE MITOCHONDRIAL OUTER MEMBRANE
CC AND ALSO THE PLASMA MEMBRANE. THE CHANNEL ALLOWS DIFFUSION OF
CC SMALL HYDROPHILIC MOLECULES; IT ADOPTS AN OPEN CONFORMATION AT LOW
CC OR ZERO MEMBRANE POTENTIAL AND A CLOSED CONFORMATION AT POTENTIALS
CC ABOVE 30-40 MV. THE OPEN STATE HAS A WEAK ANION SELECTIVITY
CC WHEREAS THE CLOSED STATE IS CATION-SELECTIVE (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE OF MITOCHONDRIA AND PLASMA
CC MEMBRANE (BY SIMILARITY).

CC -1- DOMAIN: CONSISTS MAINLY OF MEMBRANE-SPANNING SIDED BETA-SHEETS.
CC -1- SIMILARITY: BELONGS TO THE EUKARYOTIC MITOCHONDRIAL PORIN FAMILY.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: AF209725; AAF22835.1; -.
CC InterPro: IPR001925; Euk_porin.

DR Pfam: PF01459; Euk_porin.1.
DR PRINTS: PR00185; EUKARYTPORIN.

DR PROSITE: PS00558; EUKARYOTIC_PORIN; 1.
KW Outer membrane; Porin; Mitochondrion; Acetylation.

FT INIT_MEF 0 0 BY SIMILARITY.
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
FT BINDING 72 72 DICICLOHEXYLCARBODIIMIDE.
SQ SEQUENCE 282 AA: 30609 MW: 59577966f108a298 CRC64:

Query Match 38.1%; Score 40; DB 1; Length 282;
Best Local Similarity 41.2%; Pred. No. 23;
Matches 7; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 2 VSVGCVGFDPASEPDSR 18
Db 122 INLCVDVDFDIAGPSIR 138

RESULT 10
PRT_MOUSE STANDARD; PRT; 296 AA.
ID PRT_MOUSE
AC 060932;
DT 15-JUL-1999 (Rel. 38, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE VOLTAGE-DEPENDENT ANION-SELECTIVE CHANNEL PROTEIN 1 (VDAC-1) (MVDAC1)
DE (MVDAC5) (OUTER MITOCHONDRIAL MEMBRANE PROTEIN PORIN 1) (PLASMALEMAL
DE PORIN).
GN VDAC1 OR VDAC5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognath; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM MT-VDAC1).
RP TISSUE=Brain;
RC MEDLINE=96301405; PubMed=8660977;
RA Sampson M.J., Lovell R.S., Craigen W.J.;
RT "Isolation, characterization, and mapping of two mouse mitochondrial
RT voltage-dependent anion channel isoforms.";
RL Genomics 33:283-288(1996).
RN [2]
RP CHARACTERIZATION AND ALTERNATIVE SPLICING.
RP MEDLINE=20202612; PubMed=10716730;
RA Buetner R., Papoutsoglou G., Scemes E., Spray D.C., Dermietzel R.;
RT "Evidence for secretory pathway localization of a voltage-dependent
RT anion channel isoform.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:3201-3206(2000).
CC -1- FUNCTION: FORMS A CHANNEL THROUGH THE MITOCHONDRIAL OUTER MEMBRANE
CC AND ALSO THE PLASMA MEMBRANE. THE CHANNEL ALLOWS DIFFUSION OF
CC SMALL HYDROPHILIC MOLECULES: IT ADOPTS AN OPEN CONFORMATION AT POTENTIALS
CC OR ZERO MEMBRANE POTENTIAL AND A CLOSED CONFORMATION AT POTENTIALS
CC ABOVE 30-40 MV. THE OPEN STATE HAS A WEAK ANION SELECTIVITY
CC WHEREAS THE CLOSED STATE IS CATION-SELECTIVE.
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL VDAC1 (MT-VDAC1) IN OUTER
CC MEMBRANE OF MITOCHONDRIA AND PLASMALEMAL VDAC1 (PL-VDAC1) IN
CC PLASMA MEMBRANE.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: PL-VDAC1 (SHOWN HERE) AND MT-
CC VDAC1; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: HIGH LEVELS OF EXPRESSION DETECTED IN HEART,
CC KIDNEY, BRAIN, AND SKELETAL MUSCLE. NOT EXPRESSED IN TESTIS.
CC -1- DOMAIN: CONSISTS MAINLY OF THE EUKARYOTIC MITOCHONDRIAL PORIN FAMILY.
CC -1- SIMILARITY: BELONGS TO THE EUKARYOTIC MITOCHONDRIAL PORIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U30840; AAB47777.1; -
CC SWISS-2DPAGE: G60932; MOUSE.
CC MGD: MGI:106919; Vdac1.
CC InterPro: IPR001925; Euk_porin.

DR Pfam: PF01459; Euk_porin.1.
DR PRINTS: PRO0185; EUKARYTPORIN.
DR PROSITE: PS00558; EUKARYOTIC_PORIN.1.
KW Outer membrane; Porin; Mitochondrion; Alternative splicing;
KW Acetylation.
FT MOD_RES 14 14 ACETYLATION (BY SIMILARITY).
FT BINDING 86 86 DICICLOHEXYLCARBODIIMIDE (BY SIMILARITY).
FT VARSPLIC 1 13 MISSING (IN ISOFORM MT-VDAC1).
SQ SEQUENCE 296 AA: 32351 MW: C0710C1717063B52 CRC64:

Query Match 38.1%; Score 40; DB 1; Length 296;
Best Local Similarity 41.2%; Pred. No. 24;
Matches 7; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 2 VSVGCVGFDPASEPDSR 18
Db 136 INLCVDVDFDIAGPSIR 152

RESULT 11
PCK_PAPPR STANDARD; PRT; 367 AA.
ID PCK_PAPPR
AC 000869;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE PHOSPHOGLYCERATE KINASE (EC 2.7.2.3) (FRAGMENT).
DE PCK.
GN Paramecium primaurella.
OS Paramecium.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peritritida;
OC Paramecium.
OX NCBI_TaxID=5886;
RN [1]
RP SEQUENCE FROM N.A.
RA Pearlman R.E.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: ATP + 3-PHOSPHO-D-GLYCERATE = ADP +
CC 3-PHOSPHO-D-GLYCEROYL PHOSPHATE.
CC -1- PATHWAY: SECOND STEP IN THE SECOND PHASE OF GLYCOLYSIS.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE PHOSPHOGLYCERATE KINASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF001849; AAB58241.1; -
CC InterPro: IPR001576; PCK.
DR Pfam: PF00162; PCK.1.
DR PROSITE: PS00111; PGLYCERATE_KINASE; PARTIAL.
KW transferase; kinase; glycolysis.
FT NON_TER 1 1
FT NON_TER 367 367
SQ SEQUENCE 367 AA: 39876 MW: DEF64500127DE460 CRC64;

Query Match 38.1%; Score 40; DB 1; Length 367;
Best Local Similarity 42.1%; Pred. No. 30;
Matches 8; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

QY 1 PVSVCVGFDPASEPDSRH 19
Db 259 PDPFVCGTGIDASSPVALH 277

RESULT 12
KYNU_HUMAN STANDARD; PRT; 465 AA.
ID KYNU_HUMAN

AC 016719;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE KYNURENINASE (EC 3.7.1.3) (L-KYNURENINE HYDROLASE).
 GN KYN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RC TISSUE=Hepatosoma;
 RX MEDLINE=96314506; PubMed=8706755;
 RA Oberlander G., Buchli R., Malherbe P., Broger C., Lang G.,
 RA Koehler C., Lahm H.-W., Cesura A.M.;
 RT "Isolation and expression of a cDNA clone encoding human
 kynureninase.";
 RL Eur. J. Biochem. 239:460-468(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=97324088; PubMed=9180257;
 RA Toma S., Nakamura M., Tone S., Okuno E., Kido R., Breton J.,
 RA Avanzi N., Cozzi L., Spectale C., Mostardini M., Gatti S., Benatti L.;
 RT "Cloning and recombinant expression of rat and human kynureninase.";
 RL FEBS Lett. 408:5-10(1997).
 CC -1- FUNCTION: CATALYZES THE CLEAVAGE OF L-KYNURENINE (L-KYN) AND L-3-
 HYDROXYKYNURENINE (L-3OHKYN) INTO ANTHRANILIC (AA) AND 3-
 HYDROXYANTHRANILIC ACIDS (3-OHAA), RESPECTIVELY. HAS A PREFERENCE
 FOR THE L-3-HYDROXY FORM. ALSO HAS CYSTEINE-CONJUGATE-BETA-LYASE
 ACTIVITY (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: L-KYNURENINE + H(2)O = ANTHRANILATE +
 L-ALANINE.
 CC -1- CATALYTIC ACTIVITY: L-3-HYDROXYKYNURENINE + H(2)O = 3-
 HYDROXYANTHRANILATE + L-ALANINE.
 CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.
 CC -1- ENZYME REGULATION: INHIBITED BY O-METHOXYBENZYLALANINE (OMBA).
 CC -1- PATHWAY: INVOLVED IN THE BIOSYNTHESIS OF NAD COFACTORS FROM
 TRYPHOPHAN THROUGH THE KYNURENINE PATHWAY.
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN ALL TISSUES TESTED (HEART, BRAIN
 PLACENTA, LUNG, LIVER, SKELETAL MUSCLE, KIDNEY AND PANCREAS).
 CC HIGHEST LEVELS FOUND IN PLACENTA, LIVER AND LUNG. EXPRESSED IN
 ALL BRAIN REGIONS.
 CC -1- INDUCTION: INCREASED LEVELS IN SEVERAL CEREBRAL AND SYSTEMIC
 INFLAMMATORY CONDITIONS.
 CC -1- SIMILARITY: BELONGS TO THE KYNURENINASE FAMILY. SLIGHTLY RELATED
 TO CLASS-V OF PYRIDOXAL-PHOSPHATE-DEPENDENT AMINOTRANSFERASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U57721; AAC50650.1; -;
 DR MIM: 605197; -;
 DR MIM: 236800; -;
 DR Hydrolase: Pyridoxal phosphate: Acetylation.
 KM MOD_RES 1 ACETYLATION (BY SIMILARITY).
 FT BINDING 276 276 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
 SQ SEQUENCE 465 AA; 52351 MW; BDD136BE18C79EBB CRC64;

Query Match 38.1%; Score 40; DB 1; Length 465;
 Best Local Similarity 85.7%; Pred. No. 39;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Oy 5 GCFFVGD 11

DB 244 GCYVGF 250
 ||:||||
 RESULT 13
 ID ALBU_CHICK STANDARD; PRT: 615 AA.
 AC P19121;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE SERUM ALBUMIN PRECURSOR.
 GN ALB.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX Cassady A.I., Salkind C.K., Baverstock P., Wallace J.C.;
 RL Submitted (JUL-1991) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 1-28 FROM N.A.
 RX MEDLINE=83161037; PubMed=6187737;
 RA Hache R.J.G., Wiskocil R., Vasa M., Roy R.N., Lau P.C.K., Deeley R.G.;
 RT "The 5' noncoding and flanking regions of the avian very low density
 apolipoprotein II and serum albumin genes. Homologies with the egg
 white protein genes.";
 RL J. Biol. Chem. 258:4556-4564(1983).
 RN [3]
 RP SEQUENCE OF 19-30.
 RX MEDLINE=78019943; PubMed=911327;
 RA Rosen A.M., Geller D.M.;
 RT "Chicken microsomal albumin: amino terminal sequence of chicken
 prealbumin.";
 RL Biochem. Biophys. Res. Commun. 78:1060-1066(1977).
 CC -1- FUNCTION: SERUM ALBUMIN, THE MAIN PROTEIN OF PLASMA, HAS A GOOD
 BINDING CAPACITY FOR WATER, CA⁺⁺, NA⁺, K⁺, FATTY ACIDS, HORMONES,
 BILIRUBIN AND DRUGS. ITS MAIN FUNCTION IS THE REGULATION OF THE
 COLLOIDAL OSMOTIC PRESSURE OF BLOOD.
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
 CC -1- TISSUE SPECIFICITY: PLASMA.
 CC -1- DOMAIN: COMPOSED OF THREE HOMOLOGOUS DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE ALB/APP/VDB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X60688; CAA43098.1; -;
 DR EMBL: V00381; CAA23680.1; -;
 DR PIR: S15711; ABCHS.
 DR HSP: P02768; 1DOR.
 DR InterPro: IPR000264; Serum_albumin.
 DR Pfam: PF00273; transport_prot. 3.
 DR PRINTS: PRO0802; SERUMALBUMIN.
 DR SMART: SM00103; ALBUMIN; 3.
 DR PROSITE: PS00212; ALBUMIN; 3.
 KM Plasma; Metal-binding; Lipid-binding; Albumin; Repeat; Signal;
 KW Copper.
 FT SIGNAL 1 18
 FT PROPEP 19 23
 FT CHAIN 24 615
 FT REPEAT 31 206
 FT REPEAT 225 398
 FT REPEAT 417 596
 FT METAL 30 30
 COPPER (BY SIMILARITY).

```
FT DISULFID 80 89 BY SIMILARITY..
FT DISULFID 102 118 BY SIMILARITY..
FT DISULFID 117 128 BY SIMILARITY..
FT DISULFID 152 197 BY SIMILARITY..
FT DISULFID 196 205 BY SIMILARITY..
FT DISULFID 228 274 BY SIMILARITY..
FT DISULFID 273 281 BY SIMILARITY..
FT DISULFID 293 307 BY SIMILARITY..
FT DISULFID 306 317 BY SIMILARITY..
FT DISULFID 344 389 BY SIMILARITY..
FT DISULFID 388 397 BY SIMILARITY..
FT DISULFID 420 466 BY SIMILARITY..
FT DISULFID 465 476 BY SIMILARITY..
FT DISULFID 489 505 BY SIMILARITY..
FT DISULFID 504 515 BY SIMILARITY..
FT DISULFID 542 587 BY SIMILARITY..
FT DISULFID 586 595 BY SIMILARITY..
FT CARBOHYD 500 500 N-LINKED (GLCNAC.. ) (POTENTIAL).
FT CONFLICT 24 24 F -> M (IN REF. 3).
SQ SEQUENCE 615 AA; 69918 MW; E59E4B8CAEC066C6 CRC64;

Query Match 38.1%; Score 40; DB 1; Length 615;
Best Local Similarity 54.5%; Pred. No. 51;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 6 CPVGFADSEPD 16
Db 128 CPLSFVKSQPD 138

RESULT 14
PEX8_PICAN STANDARD; PRT; 650 AA.
ID PEX8_PICAN
AC 000925;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE PEROXISOMAL MATRIX PROTEIN PERI PRECURSOR (PEROXIN-8).
GN PEX8 OR PERI.
OS Pichia angusta (Yeast) (Hansenula polymorpha).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Pichia.
OC NCBI_TaxID=4905;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CBS 4732;
RX MEDLINE=95050945; PubMed=7962056;
RX Waterham H.R., Titorenko V.I., Haima P., Gregg J.M., Harder W.,
RX Veenhuis M.;
RX "The Hansenula polymorpha Per1 gene is essential for peroxisome
RX biogenesis and encodes a peroxisomal matrix protein with both
RX carboxy- and amino-terminal targeting signals.";
RL J. Cell Biol. 127:737-749(1994)
CC -!- FUNCTION: ESSENTIAL FOR PEROXISOME BIOGENESIS. MAY PLAY A ROLE IN
CC TRIGGERING THE PROTEIN IMPORT COMPETENCE OF INDIVIDUAL
CC SUBCELLULAR LOCATIONS. IT MAY INTERACT WITH PEX8 (PEX10).
CC -!- SUBCELLULAR LOCATION: PEROXISOMAL; MATRIX.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z30206; CAA82928.1; -
CC DR PROSITE; PS00342; MICROBODIES_CTER; 1.
CC KW Peroxisome; Transit peptide.
CC FT TRANSIT 1 650 MICROBODY.
CC FT CHAIN 1 650 PEROXISOMAL MATRIX PROTEIN PERI.
CC FT SITE 648 650 MICROBODY TARGETING SIGNAL (POTENTIAL).
```

```
SQ SEQUENCE 650 AA; 74122 MW; A9AC534204F50C7D CRC64;

Query Match 38.1%; Score 40; DB 1; Length 650;
Best Local Similarity 36.8%; Pred. No. 54;
Matches 7; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 1 PVSVCVFADSEPD 19
Db 145 PVAAGCLVDSRNDYDR 163

RESULT 15
AMO_KLEAE STANDARD; PRT; 755 AA.
ID AMO_KLEAE
AC P49250;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE COPPER AMINE OXIDASE PRECURSOR (EC 1.4.3.6) (MONAMINE OXIDASE)
DE (TYRAMINE OXIDASE).
DE MAOA OR TYNA.
OS Klebsiella aerogenes.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
OC NCBI_TaxID=28451;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-41.
RX STRAIN=M70;
RX MEDLINE=92210491; PubMed=1556068;
RX Sugino H., Sasaki M., Azakami H., Yamashita M., Murooka Y.;
RX "A monamine-regulated Klebsiella aerogenes operon containing the
RX monamine oxidase structural gene (maoa) and the maoc gene.";
RX J. Bacteriol. 174:2485-2492(1992).
CC -!- FUNCTION: ACTIVE ON TYRAMINE, TRYPTAMINE, BETA-PHENETHYLAMINE AND
CC DOPAMINE.
CC -!- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) = RCHO + NH(3) +
CC H(2)O(2).
CC -!- COFACTOR: COPPER AND TOPAQUINONE.
CC -!- SUBCELLULAR LOCATION: PERIPLASMIC.
CC -!- INDUCTION: BY TYRAMINE AND CATECHOLAMINES.
CC -!- PTM: THE TOPA QUINONE IS GENERATED BY COPPER-DEPENDENT
CC AUTOXIDATION OF A SPECIFIC TYROSYL RESIDUE (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D10208; BAA01060.1; -
CC DR HSSP; P46883; 10AC.
CC DR InterPro; IPR000269; Copper_amine_oxidase.
CC DR Pfam; PF01179; Cu_amine_oxid_1.
CC DR Pfam; PF02727; Cu_amine_oxid2_1.
CC DR Pfam; PF02728; Cu_amine_oxid3_1.
CC DR PROSITE; PS01164; COPPER_AMINE_OXID_1; 1.
CC DR PROSITE; PS01165; COPPER_AMINE_OXID_2; 1.
CC KW Oxidoreductase; Copper; TPQ; Periplasmic; Signal.
CC FT SIGNAL 1 30
CC FT CHAIN 31 755 COPPER AMINE OXIDASE.
CC FT ACT SITE 413 413 CATALYTIC BASE (BY SIMILARITY).
CC FT BINDING 496 496 TOPAQUINONE (BY SIMILARITY).
CC FT METAL 554 554 COPPER (POTENTIAL).
CC FT METAL 556 556 COPPER (POTENTIAL).
CC FT METAL 719 719 COPPER (POTENTIAL).
CC FT METAL 719 719 COPPER (POTENTIAL).
SQ SEQUENCE 755 AA; 83576 MW; 7B5552283CD93EFF CRC64;

Query Match 38.1%; Score 40; DB 1; Length 755;
```


Best Local Similarity 44.48; Pred. No. 63;
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Oy 1 PVSVCFCVGFDASEPPSR 18
|::||| | | : |::|
Db 254 PLTVGFFDGKDGLOQDAR 271

Search completed: November 30, 2001, 09:40:55
Job time: 34 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 30, 2001, 09:44:52 ; Search time 67.45 Seconds
(without alignments)
41.203 Million cell updates/sec

Title: US-08-846-234-2
Perfect score: 105
Sequence: 1 PVSVCFCVGFDAEPDSRH 19

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP:archaea:*
2: SP:bacteria:*
3: SP:fungi:*
4: SP:human:*
5: SP_invertebrate:*
6: SP_mammal:*
7: SP_mhc:*
8: SP_organelle:*
9: SP_phage:*
10: SP_plant:*
11: SP_todent:*
12: SP_virus:*
13: SP_vertebrate:*
14: SP_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	105	100.0	784	10 Q9ZT62	Q9ZT62 cucumis sat
2	50	47.6	779	10 Q9M4M7	Q9M4M7 persica amer
3	47	44.8	295	10 Q8Z796	Q8Z796 arabidopsis
4	47	44.8	295	10 Q9FZ85	Q9FZ85 arabidopsis
5	46.5	44.3	783	10 Q9FND9	Q9FND9 arabidopsis
6	45	42.9	757	13 Q9YHD2	Q9YHD2 gallus gall
7	44	41.9	202	11 Q9QVM8	Q9QVM8 rattus sp.
8	44	41.9	301	2 Q43942	Q43942 mycobacteri
9	44	41.9	421	5 Q9VWQ2	Q9VWQ2 dirosophila
10	44	41.9	498	10 Q9LZ54	Q9LZ54 arabidopsis
11	44	41.9	508	10 Q9C5N0	Q9C5N0 arabidopsis
12	44	41.9	650	11 Q63709	Q63709 rattus norv
13	44	41.9	742	10 Q9AV10	Q9AV10 oryza sativ
14	43	41.0	167	9 Q9KCH9	Q9KCH9 streptococc
15	43	41.0	180	5 Q9U249	Q9U249 caenorhabd1
16	43	41.0	236	10 Q9XEQ8	Q9XEQ8 sorghum bic
17	43	41.0	313	5 Q9GZFI	Q9GZFI caenorhabd1
18	43	41.0	441	10 Q23053	Q23053 arabidopsis
19	43	41.0	447	10 Q80885	Q80885 arabidopsis

20	43	41.0	464	11 Q9CXF0	Q9CXF0 mus musculu
21	42.5	40.5	546	2 Q31813	Q31813 bacillus su
22	42	40.0	169	1 Q9YB34	Q9YB34 acetyprum p
23	42	40.0	240	2 Q9KBE5	Q9KBE5 bacillus ha
24	42	40.0	257	2 Q51720	Q51720 propionibac
25	42	40.0	311	5 Q44789	Q44789 caenorhabd1
26	42	40.0	314	5 Q9XW86	Q9XW86 caenorhabd1
27	42	40.0	376	10 Q9FVH5	Q9FVH5 prunus arme
28	42	40.0	393	2 Q9LCH6	Q9LCH6 streptomyc
29	42	40.0	419	2 Q08410	Q08410 shewanella
30	42	40.0	439	2 Q52785	Q52785 actinobact
31	42	40.0	511	5 Q9XUS4	Q9XUS4 caenorhabd1
32	42	40.0	648	10 Q9M824	Q9M824 arabidopsis
33	42	40.0	653	2 Q9A2E5	Q9A2E5 caulobacter
34	42	40.0	1132	10 Q9AYH7	Q9AYH7 oryza sativ
35	41.5	39.5	101	2 Q9X4M7	Q9X4M7 lactobacill
36	41.5	39.5	219	12 Q39302	Q39302 equine herp
37	41.5	39.5	222	10 Q9FMB0	Q9FMB0 arabidopsis
38	41.5	39.5	360	3 Q9P8B3	Q9P8B3 agaricus bl
39	41	39.0	125	2 Q9JS17	Q9JS17 neisseria m
40	41	39.0	251	2 Q9X3Y6	Q9X3Y6 zymomonas m
41	41	39.0	283	6 Q9M216	Q9M216 sus scrofa
42	41	39.0	314	5 Q23017	Q23017 caenorhabd1
43	41	39.0	511	2 Q9A4C3	Q9A4C3 caulobacter
44	41	39.0	552	5 Q20329	Q20329 caenorhabd1
45	41	39.0	739	12 Q91WA0	Q91WA0 pothos late

ALIGNMENTS

RESULT 1	Q9ZT62	PRELIMINARY:	PRT:	784 AA.
AC	Q9ZT62:			
DT	01-MAY-1999 (TREMBlrel. 10, Created)			
DT	01-MAY-1999 (TREMBlrel. 10, last sequence update)			
DT	01-MAY-2000 (TREMBlrel. 13, last annotation update)			
DE	RAFTINASE SYNTHASE (EC 2.4.1.82).			
GN	RF5.			
OS	Cucumis sativus (Cucumber).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	eurosid1; Cucurbitales; Cucurbitaceae; Cucumis.			
OX	NCBI_TaxID=3659;			
RM	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=LEAF;			
RA	Ohsumi C., Nozaki J., Kida T.;			
RL	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF073744; AAD02832.1; ..			
KW	Transferase; Glycosyltransferase.			
SQ	SEQUENCE 784 AA; 86919 MW; 3B06A491F0908933 CRC64;			
Query Match	100.0%;	Score 105;	DB 10;	Length 784;
Best Local Similarity	100.0%;	Pred. No. 1.1e-08;		
Matches 19;	Conservative	0;	Mismatches 0;	Indels 0;
			Gaps 0;	
QY	1 PVSVCFCVGFDAEPDSRH 19			
DB	61 PVSVCFCVGFDAEPDSRH 79			
RESULT 2	Q9M4M7	PRELIMINARY:	PRT:	779 AA.
ID	Q9M4M7:			
AC	Q9M4M7:			
DT	01-OCT-2000 (TREMBlrel. 15, Created)			
DT	01-OCT-2000 (TREMBlrel. 15, last sequence update)			
DT	01-OCT-2000 (TREMBlrel. 15, last annotation update)			
DE	POTATIVE SEED IMBIBITION PROTEIN.			
GN	SIP.			

OS Persea americana (Avocado).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Laurales; Lauraceae; Persea.
OX NCBI_TaxID=3435;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. HASS; TISSUE=FRUIT MESOCARP;
RA Zamorano J.P., Evans A.D., Dopico B., Lowe A.L., Wilson I.D.,
RA Merodio C., Grierson D.;
RT "Isolation and characterization of cDNAs for mRNAs regulated during
RT cold storage of avocado (Persea americana Mill.) fruit."
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ133148; CAB77245.1; -
SQ SEQUENCE 779 AA; 85368 MW; C3A8B43160316785 CRC64;

Query Match 47.6%; Score 50; DB 10; Length 779;
Best Local Similarity 60.0%; Pred. No. 9.1;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
OY 5 GCYVGFDPASEPDSRH 19
|||:| |||:| |
Db 42 GCFLGATASESESIH 56

RESULT 3
ID 082796 PRELIMINARY; PRT; 295 AA.
AC 082796;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE 3-PHOSPHOSERIN PHOSPHATASE.
GN PSP.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
[1]
RP SEQUENCE FROM N.A.
RA HO C., Noji M., Salto K.;
RT "Molecular cloning and characterization of PSP."
RT Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RL EMBL; AB018409; BAA33807.1; -
DR EMBL; AB018408; BAA33806.1; -
DR InterPro: IPR001454; Hydrolase.
DR Pfam: PF00702; Hydrolase; 1.
SQ SEQUENCE 295 AA; 32302 MW; F14C95E636E754E CRC64;

Query Match 44.8%; Score 47; DB 10; Length 295;
Best Local Similarity 64.3%; Pred. No. 10;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
OY 5 GCYVGFDPASEPDSRH 18
| | : | | | : | | | |
Db 210 GEFLLGFDENEPTSR 223

RESULT 4
ID 09F285 PRELIMINARY; PRT; 295 AA.
AC 09F285;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE 3-PHOSPHOSERINE PHOSPHATASE.
GN P26116.2.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;

OX NCBI_TaxID=3702;
[1]
RP SEQUENCE FROM N.A.
RA Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
RA Altati H., Nguyen M., Lam B., Southwick A., Miranda M., Brooks S.,
RA Buehler E., Chao Q., Chin C., Chlou J., Choi E., Gonzalez A.,
RA Hong B., Johnson-Hopson C., Khan S., Kim C., Koo T., Lee J.M.,
RA Lenz C., Liu A., Liu S., Mukharisky N., Pham P., Sakano H., Shin P.,
RA Toriumi M., Vayenberg M., Yu G., Ecker J., Theologis A., Davis R.W.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC026238; AAF98410.1; -
SQ SEQUENCE 295 AA; 32318 MW; F14C95E636E7745E CRC64;

Query Match 44.8%; Score 47; DB 10; Length 295;
Best Local Similarity 64.3%; Pred. No. 10;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
OY 5 GCYVGFDPASEPDSRH 18
| | : | | | : | | | |
Db 210 GEFLLGFDENEPTSR 223

RESULT 5
ID 09FND9 PRELIMINARY; PRT; 783 AA.
AC 09FND9;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE RAFFINOSE SYNTHASE PROTEIN.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RX MEDLINE=98069011; PubMed=9405937;
RA Kotani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. II.
RT Sequence features of the regions of 1,044,062 bp covered by thirteen
RT physically assigned P1 clones."
RL DNA Res. 4:291-300(1997).
DR EMBL; AB006702; BAB11595.1; -
SQ SEQUENCE 783 AA; 86237 MW; 3C37D1D7871888AC CRC64;

Query Match 44.3%; Score 46.5; DB 10; Length 783;
Best Local Similarity 52.6%; Pred. No. 34;
Matches 10; Conservative 2; Mismatches 6; Indels 1; Gaps 1;
OY 2 VSVGCFVGFDA-SEPDSSRH 19
| | : | | | : | | | |
Db 62 VSAGSFGFNDGEPKSHH 80

RESULT 6
ID 09YHD2 PRELIMINARY; PRT; 757 AA.
AC 09YHD2;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE NUCLEAR CALMODULIN-BINDING PROTEIN (FRAGMENT).
GN URP.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;

RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN:
 RA Lodge A.P., Walsh A., McNamee C.J., Moss D.J.;
 RT "Identification of Churp, a Nuclear Calmodulin-Binding Protein Related
 to hNRNP-U.";
 RL Submitted (Oct-1998) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF098788; AAC69888.1; -
 DR InterPro: IPR003877; SPRY.
 DR InterPro: IPR003878; SPRY_Domain.
 DR Pfam: PF00622; SPRY; 1.
 DR SMART: SM00449; SPRY; 1.
 FT NON-TER
 SQ SEQUENCE 757 AA; 84178 MW; 7FA04ED9E1776D1D CRC64;

Query Match 42.9%; Score 45; DB 13; Length 757;
 Best Local Similarity 53.8%; Pred. No. 57;
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 4 VCGCFGVDFASEPD 16
 DB 342 ICGFADEFASEE 354

RESULT 7
 ID 090VW8 PRELIMINARY; PRT; 202 AA.

AC 090VW8;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE FIBROBLAST GROWTH FACTOR RECEPTOR 4, FGFR-4.
 OS Rattus sp.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10118;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94293355; PubMed=8021968;
 RA Yazaki N., Hosaki Y., Kawabata K., Miyake A., Minami M., Satoh M.,
 RA Ohta M., Kawasaki T., Itoh N.;
 RT "Differential expression patterns of mRNAs for members of the
 RT fibroblast growth factor receptor family, FGFR-1-FGFR-4, in rat
 RT brain.";
 RL J. Neurosci. Res. 37:445-452(1994).
 DR HSP: P06239; 31CK.
 DR InterPro: IPR000719; Euk_pkinase.
 DR Pfam: PF00069; pkinase; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 KW ATP-binding; Transferase.
 SQ SEQUENCE 202 AA; 21986 MW; 8F87690566BFA31 CRC64;

Query Match 41.9%; Score 44; DB 11; Length 202;
 Best Local Similarity 37.5%; Pred. No. 21;
 Matches 9; Conservative 3; Mismatches 4; Indels 8; Gaps 1;

OY 1 PVSVCGF-----VGFDFASEPD 16
 DB 149 PLCEGCFGVFAALGMSRSD 172

RESULT 8
 ID 049942 PRELIMINARY; PRT; 301 AA.

AC 049942;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE HYPOTHEICAL 33.9 KDA PROTEIN L518_F3_81.
 GN ML2346 OR L518_F3_81 OR MCB2407.04.

OS Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1769;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-TN;
 RA Cole S.T., Elgimeier K., Parkhill J., James K.D., Thomson N.R.,
 RA Wheeler P.R., Honore N., Ganier T., Churcher C., Harris D.,
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R.M., Devlin K., Duthoy S., Fellwell T., Fraser A., Hamlin N.,
 RA Holtroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
 RA Murphy L., Oliver, Quail M.A., Rajandream M.A., Rutherford K.M.,
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
 RA Barrell B.G.;
 RT "Massive gene decay in the leprosy bacillus.";
 RL Nature 409:1007-1011(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Robison K., Smith D.R.;
 RL Submitted (MAR-1994) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Seeger K., Harris D.;
 RL Submitted (MAY-1998) to the EMBL/Genbank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (MAY-1998) to the EMBL/Genbank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93188700; PubMed=8446027;
 RA Elgimeier K., Honore N., Woods S.A., Caudron B., Cole S.T.;
 RL Mol. Microbiol. 7:197-206(1993).
 DR EMBL: AL583925; CAC31862.1; -
 DR EMBL: U00023; AAA17366.1; -
 DR EMBL: AL023596; CAA19144.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 301 AA; 33939 MW; CF11EF2EEAE7A0AD CRC64;

Query Match 41.9%; Score 44; DB 2; Length 301;
 Best Local Similarity 57.1%; Pred. No. 32;
 Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 PVSVCGFVDFASE 14
 DB 200 PAREGCFGVDFTE 213

RESULT 9
 ID 09VVO2 PRELIMINARY; PRT; 421 AA.

AC 09VVO2;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
 DE CG7271 PROTEIN.
 GN CG7271.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Gelinkker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borikova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahle C., Daveport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshireli A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,
RA Williams S.M., Woodard T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003520; AAF49258.1; -;
DR FLYBASE: FBgn0036791; CG7271.
SQ SEQUENCE 421 AA; 47361 MW; F0519C00EFAA569C CRC64;

Query Match 41.9%; Score 44; DB 5; Length 421;
Best Local Similarity 50.0%; Pred. No. 45;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 6 GCFVGFADSEPDSSRH 19
ID 1: 11 11 11 11
DB 36 CYSSDADEPDSSRH 49
RESULT 10
Q91Z54 PRELIMINARY; PRT; 498 AA.
AC Q91Z54;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HYPOTHEICAL 55.5 KDA PROTEIN.
GN T22P11_70.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA Bancroft I., Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL162971; CAB85985.1; -;
DR InterPro: IPR002068; Crystallin_HSP20.

DR InterPro: IPR000834; Zn_Carboxypept.
DR PROSITE: PS00133; CARBOXYPEPT_2N_2; UNKNOWN_1.
DR PROSITE: PS01031; HSP20; 1.
KW Hypothetical protein.
SQ SEQUENCE 498 AA; 55530 MW; 0791437EB8817BE2 CRC64;

Query Match 41.9%; Score 44; DB 10; Length 498;
Best Local Similarity 70.0%; Pred. No. 54;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 5 GCFVGFADSE 14
ID 132 GCFVGFADSKSD 141
DB 132 GCFVGFADSKSD 141

RESULT 11
Q9C5N0 PRELIMINARY; PRT; 508 AA.
AC Q9C5N0;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HYPOTHEICAL 56.6 KDA PROTEIN.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
RA Seki M., Shinn P., Southwick A., Shinzaki K., Davis R.W., Ecker J.R.,
RA Theologis A.;
RT "Full Length cDNA of gene T22P11.70/At5g02480 (GI:7413637).";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF360137; AAK25847.1; -;
DR InterPro: IPR002068; Crystallin_HSP20.
DR InterPro: IPR000834; Zn_Carboxypept.
DR PROSITE: PS00133; CARBOXYPEPT_2N_2; UNKNOWN_1.
DR PROSITE: PS01031; HSP20; 1.
KW Hypothetical protein.
SQ SEQUENCE 508 AA; 56561 MW; 8A4406B84E15B33C CRC64;

Query Match 41.9%; Score 44; DB 10; Length 508;
Best Local Similarity 70.0%; Pred. No. 55;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 5 GCFVGFADSE 14
ID 142 GCFVGFADSKSD 151
DB 142 GCFVGFADSKSD 151
RESULT 12
Q63709 PRELIMINARY; PRT; 650 AA.
AC Q63709;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE FIBROBLAST GROWTH FACTOR RECEPTOR SUBTYPE 4.
GN FGFR4.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.

RX MEDLINE-93013049; PubMed-1398143;
RA Horlick R.A., Slack S.L., Cooke G.M.;
RT "Cloning, expression and tissue distribution of the gene encoding rat
RT fibroblast growth factor receptor subtype 4.";
RL Gene 120:291-295(1992).
CC -1- SIMILARITY: TO IMMUGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
DOMAIN.
DR EMBL: M91599; AAA1157.1; -.
DR HSSP: P06239; 3ICK.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR003598; Iq_c2.
DR InterPro: IPR003006; Iq_MHC.
DR InterPro: IPR01245; Tyr_kin.
DR Pfam: PF00047; Iq; 2.
DR Pfam: PF00069; pkinase; 1.
DR PRINTS: PR00109; TYRKINASE.
DR SMART: SM00408; IGC2; 2.
DR SMART: SM00219; TYRKC; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 650 AA; 71858 MW; 78A7F448FB973867 CRC64;

Query Match 41.9%; Score 44; DB 11; Length 650;
Best Local Similarity 37.5%; Pred. No. 71;

Matches 9; Conservative 3; Mismatches 4; Indels 8; Gaps 1;

OY 1 PVSVCGF-----VGFDPSEP 16
ID 1: 111 : 111 : 111 :
DB 320 PLGEGCGGVCAEALGMDSSRD 343

RESULT 13
O9AV10 PRELIMINARY; PRT; 742 AA.
AC O9AV10;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PUTATIVE OXYSTEROL-BINDING PROTEIN.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxId=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV, NIPPONBARE;
RA Buell C.R., Yuan Q., Ouyang S., Moffat K.S., Hill J.N., Burr P.C.,
RA Hsiao J., Zisman V., Pal G., Bowman C.L., Fujii C.Y., Vanaken S.E.,
RA Bowman C.L., Creven B., Uterback T.R., Khalak H., Feldblum T.V.,
RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.;
RT "Oryza sativa chromosome 10 BAC OSJNB001411 genomic sequence.";
RT Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC037426; AKK15443.1; -.
DR EMBL: AC037426; AKK15443.1; -.
SQ SEQUENCE 742 AA; 82345 MW; B8EEF530901BD055 CRC64;

Query Match 41.9%; Score 44; DB 10; Length 742;
Best Local Similarity 47.1%; Pred. No. 82;

Matches 8; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

OY 3 SVGCVGFDPASEPDSRH 19
ID 1: 1 1 1 1 1 : 1 1 1 : 1 1 1 :
DB 590 SMSCIIGDDASKVNSRN 606

RESULT 14
O9MCH9 PRELIMINARY; PRT; 167 AA.
ID O9MCH9
AC O9MCH9;

DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PUTATIVE DNA BINDING PROTEIN.
OS Streptococcus thermophilus bacteriophage Sf11.
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae.
OX NCBI_TaxId=78541;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-99412383; PubMed-10482618;
RA Lucchini S., Desiere F., Brussow H.;
RT "Comparative genomics of Streptococcus thermophilus phage species
RT supports a modular evolution theory.";
RL J. Virol. 73:8647-8656(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Desiere F., Lucchini S., Brussow H.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF158600; AAF63067.1; -.
DR InterPro: IPR01387; HTH_3.
DR Pfam: PF01381; HTH_3; 1.
DR SMART: SM00530; HTH_XRE; 1.
SQ SEQUENCE 167 AA; 18893 MW; 0E9AA9C765BE014E CRC64;

Query Match 41.0%; Score 43; DB 9; Length 167;
Best Local Similarity 56.2%; Pred. No. 25;

Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 2 VSVGCVGFDPASEPDS 17
ID 111 : 111 : 111 :
DB 53 VSVGYLLGLDTTENDS 68

RESULT 15
O9U249 PRELIMINARY; PRT; 180 AA.
AC O9U249;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE Y53H1B.4 PROTEIN.
OS Y53H1B.4.
GN Caenorhabditis elegans.
OC Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Felodierinae; Caenorhabditis.
OX NCBI_TaxId=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA White S.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-99069613; PubMed-9851916;
RA none;
RT "Genome sequence of the nematode C. elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL: AL132851; CAB60412.1; -.
SQ SEQUENCE 180 AA; 21484 MW; E967D88655A1D300 CRC64;

Query Match 41.0%; Score 43; DB 5; Length 180;
Best Local Similarity 50.0%; Pred. No. 27;

Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 6 CFVGFDPASEPDSRH 19
ID 11 : 1 : 1 1 1 :
DB 117 CFLOFETISPKSRH 130

Search completed: November 30, 2001, 09:44:54

Job time: 273 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: November 30, 2001, 09:43:37 ; Search time 34.66 Seconds
(without alignments)
12.336 Million cell updates/sec

Title: US-08-846-234-2
105
Sequence: 1 PVSVCFCVGFDAEPDSRH 19

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2_6/prodata/2/1aa/5A.COMB.pep: *
2: /cgn2_6/prodata/2/1aa/6A.COMB.pep: *
3: /cgn2_6/prodata/2/1aa/6A.COMB.pep: *
4: /cgn2_6/prodata/2/1aa/6B.COMB.pep: *
5: /cgn2_6/prodata/2/1aa/6C.COMB.pep: *
6: /cgn2_6/prodata/2/1aa/6C.COMB.pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	105	100.0	19	US-08-846-234-2	Sequence 2, Appl1
2	105	100.0	784	US-08-846-234-5	Sequence 5, Appl1
3	42	40.0	22	US-08-805-223-460	Sequence 460, App
4	41	39.0	212	US-08-861-774E-22	Sequence 22, Appl
5	41	39.0	212	US-08-861-774E-34	Sequence 34, Appl
6	40	38.1	291	US-09-105-697-4	Sequence 4, Appl1
7	40	38.1	291	US-09-105-697-5	Sequence 5, Appl1
8	40	38.1	291	US-09-105-697-6	Sequence 6, Appl1
9	40	38.1	291	US-07-977-434-8	Sequence 8, Appl1
10	40	38.1	834	US-07-977-434-10	Sequence 10, Appl
11	40	38.1	834	US-08-073-384C-6	Sequence 6, Appl1
12	40	38.1	834	US-08-254-359A-6	Sequence 6, Appl1
13	40	38.1	834	US-08-384-490-31	Sequence 31, Appl
14	40	38.1	834	US-08-483-043-6	Sequence 6, Appl1
15	40	38.1	834	US-08-459-383-31	Sequence 31, Appl
16	40	38.1	834	US-08-458-819-8	Sequence 8, Appl1
17	40	38.1	834	US-08-458-819-10	Sequence 10, Appl
18	40	38.1	834	US-08-481-238-6	Sequence 6, Appl1
19	40	38.1	834	US-08-471-066B-6	Sequence 6, Appl1
20	40	38.1	834	US-08-484-956-6	Sequence 6, Appl1
21	40	38.1	834	US-08-757-653-6	Sequence 6, Appl1
22	40	38.1	834	US-08-599-491-6	Sequence 6, Appl1
23	40	38.1	834	US-08-756-386-6	Sequence 6, Appl1
24	40	38.1	834	US-08-823-516-6	Sequence 6, Appl1
25	40	38.1	834	US-08-682-853A-6	Sequence 6, Appl1
26	40	38.1	834	US-08-759-038-6	Sequence 6, Appl1
27	40	38.1	834	US-08-758-314-6	Sequence 6, Appl1

28	40	38.1	834	5	PCT-US91-07035-8	Sequence 8, Appl1
29	40	38.1	834	5	PCT-US91-07035-10	Sequence 10, Appl
30	40	38.1	834	5	PCT-US95-14418-2	Sequence 2, Appl1
31	40	38.1	834	5	PCT-US95-15327-2	Sequence 2, Appl1
32	39.5	37.6	435	4	US-08-338-530A-2	Sequence 2, Appl1
33	39.5	37.6	435	4	US-09-267-384-2	Sequence 2, Appl1
34	39	37.1	461	4	US-07-857-224B-39	Sequence 39, Appl
35	39	37.1	481	4	US-08-843-659-4	Sequence 4, Appl1
36	39	37.1	560	5	PCT-US95-14418-5	Sequence 5, Appl1
37	39	37.1	560	5	PCT-US95-15327-5	Sequence 5, Appl1
38	39	37.1	597	5	PCT-US95-14418-4	Sequence 4, Appl1
39	39	37.1	597	5	PCT-US95-15327-4	Sequence 4, Appl1
40	39	37.1	833	1	US-08-073-384C-8	Sequence 8, Appl1
41	39	37.1	833	1	US-08-254-359A-8	Sequence 8, Appl1
42	39	37.1	833	1	US-08-483-043-8	Sequence 8, Appl1
43	39	37.1	833	1	US-08-481-238-8	Sequence 8, Appl1
44	39	37.1	833	2	US-08-471-066B-8	Sequence 8, Appl1
45	39	37.1	833	2	US-08-484-956-8	Sequence 8, Appl1

ALIGNMENTS

RESULT 1
US-08-846-234-2
; Sequence 2, Application US/08846234
; Patent No. 6166292
; GENERAL INFORMATION:
; APPLICANT: OSUMI Chieko
; APPLICANT: NOZAKI Jinshi
; APPLICANT: KIDA Takao
; TITLE OF INVENTION: RAFFINOSE SYNTHASE GENE, METHOD FOR
; TITLE OF INVENTION: PRODUCING RAFFINOSE, AND TRANSGENIC PLANT
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSPALT, P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,234
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)-413-3000
; TELEFAX: (703)-413-2220
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: Internal
; US-08-846-234-2

Query Match 100.0%; Score 105; DB 4; Length 19;
Best Local Similarity 100.0%; Pred No. 1.7e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PVSVCFCVGFDAEPDSRH 19
Db 1 PVSVCFCVGFDAEPDSRH 19

```
RESULT 2
US-08-846-234-5
; Sequence 5, Application US/08846234
; Patent No. 6166292
; GENERAL INFORMATION:
; APPLICANT: OSUMI Chieko
; APPLICANT: KIDA Takao
; TITLE OF INVENTION: RAFFINOSE SYNTHASE GENE, METHOD FOR
; TITLE OF INVENTION: PRODUCING RAFFINOSE, AND TRANSGENIC PLANT
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIYAK, MCCLELLAND, MAIER & NEUSTADT, P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,234
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)-413-3000
; TELEFAX: (703)-413-2220
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 784 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-846-234-5

Query Match 100.0%; Score 105; DB 4; Length 784;
Best Local Similarity 100.0%; Pred. No. 1e-08;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PVSVCCEVGFDPDSEDSRH 19
Db 61 PVSVCCEVGFDPDSEDSRH 79

RESULT 3
US-08-905-223-460
; Sequence 460, Application US/08905223
; Patent No. 6222029
; GENERAL INFORMATION:
; APPLICANT: Edwards, Jean-Baptiste D.
; APPLICANT: Duclert, Aymeric
; APPLICANT: Lacroix, Bruno
; TITLE OF INVENTION: 5' ESTS FOR SECRETED PROTEINS
; NUMBER OF SEQUENCES: 503
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 501 West Broadway
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-3505
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Win95
; SOFTWARE: Word
```

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/905,223
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned A.
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 460:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 22 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; ORIGINAL SOURCE:
; ORGANISM: Homo Sapiens
; TISSUE TYPE: Brain
; FEATURE:
; NAME/KEY: sig-peptide
; LOCATION: -17..-1
; IDENTIFICATION METHOD: Von Heijne matrix
; OTHER INFORMATION: score 5.2
; OTHER INFORMATION: seq IMCLIGIKANASS/ET
; US-08-905-223-460

Query Match 40.0%; Score 42; DB 4; Length 22;
Best Local Similarity 36.8%; Pred. No. 1.8;
Matches 7; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 1 PVSVCCEVGFDPDSEDSRH 19
Db 2 PVSIMCLIGIKANASSSETH 20

RESULT 4
US-08-861-774E-22
; Sequence 22, Application US/08861774E
; Patent No. 6297007
; GENERAL INFORMATION:
; APPLICANT: Waters, Barbara
; APPLICANT: Miao, Vivian
; APPLICANT: Ho, Yap
; TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
; TITLE OF INVENTION: BIOACTIVE MOLECULES
; FILE REFERENCE: 9993-006
; CURRENT APPLICATION NUMBER: US/08/861,774E
; CURRENT FILING DATE: 1997-05-22
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 22
; LENGTH: 212
; TYPE: PRT
; ORGANISM: Xanthoparmelia cumberlandia
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (136)
; OTHER INFORMATION: Xaa-unknown amino acid
; US-08-861-774E-22

Query Match 39.0%; Score 41; DB 4; Length 212;
Best Local Similarity 43.8%; Pred. No. 30;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 PVSVCCEVGFDPDSEPD 16
Db 22 PEDVCYIGACATDYD 37
```

```
RESULT 5
US-08-861-774E-34
; Sequence 34, Application US/08861774E
; Patent No. 6297007
; GENERAL INFORMATION:
; APPLICANT: Waters, Barbara
; APPLICANT: Miao, Vivian
; APPLICANT: Ho, Yap
; APPLICANT: Tong, Seow
; TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
; FILE REFERENCE: 9993-006
; CURRENT APPLICATION NUMBER: US/08/861,774E
; CURRENT FILING DATE: 1997-05-22
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 34
; LENGTH: 212
; TYPE: PRT
; ORGANISM: Leptogium corniculatum
US-08-861-774E-34

Query Match          39.0%; Score 41; DB 4; Length 212;
Best Local Similarity 43.8%; Pred. No. 30;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 PVSVCFCVCFDASEPD 16
1 |||:| |::|
Db 22 PEDVGCYICACATDYD 37

RESULT 6
US-09-105-697-4
; Sequence 4, Application US/09105697
; Patent No. 6228628
; GENERAL INFORMATION:
; APPLICANT: Gelfand Ph.D., David H.
; APPLICANT: Reichert, Fred L.
; TITLE OF INVENTION: MUTANT CHIMERIC DNA POLYMERASE
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Roche Molecular Systems
; STREET: 1080 U.S. Highway 202
; CITY: Branchburg
; STATE: New Jersey
; COUNTRY: United States
; ZIP: 08876
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/105,697
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Petry Ph.D., Douglas A.
; REGISTRATION NUMBER: 35321
; REFERENCE/DOCKET NUMBER: 1043
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510)814-2974
; TELEFAX: (510)814-2977
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 291 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-105-697-4
```

```
Query Match          38.1%; Score 40; DB 4; Length 291;
Best Local Similarity 61.5%; Pred. No. 62;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 7 FVGFDAEPDSRH 19
1 ||| | | |
Db 64 FVYFDKAPSFRR 76

RESULT 7
US-09-105-697-5
; Sequence 5, Application US/09105697
; Patent No. 6228628
; GENERAL INFORMATION:
; APPLICANT: Gelfand Ph.D., David H.
; APPLICANT: Reichert, Fred L.
; TITLE OF INVENTION: MUTANT CHIMERIC DNA POLYMERASE
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Roche Molecular Systems
; STREET: 1080 U.S. Highway 202
; CITY: Branchburg
; STATE: New Jersey
; COUNTRY: United States
; ZIP: 08876
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/105,697
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Petry Ph.D., Douglas A.
; REGISTRATION NUMBER: 35321
; REFERENCE/DOCKET NUMBER: 1043
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510)814-2974
; TELEFAX: (510)814-2977
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 291 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-105-697-5

Query Match          38.1%; Score 40; DB 4; Length 291;
Best Local Similarity 61.5%; Pred. No. 62;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 7 FVGFDAEPDSRH 19
1 ||| | | |
Db 64 FVYFDKAPSFRR 76

RESULT 8
US-09-105-697-6
; Sequence 6, Application US/09105697
; Patent No. 6228628
; GENERAL INFORMATION:
; APPLICANT: Gelfand Ph.D., David H.
; APPLICANT: Reichert, Fred L.
; TITLE OF INVENTION: MUTANT CHIMERIC DNA POLYMERASE
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Roche Molecular Systems
US-09-105-697-6
```

STREET: 1080 U.S. Highway 202
CITY: Branchburg
STATE: New Jersey
COUNTRY: United States
ZIP: 08876
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/105,697
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Petry Ph.D., Douglas A.
REGISTRATION NUMBER: 35321
REFERENCE/DOCKET NUMBER: 1043
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510)814-2974
TELEFAX: (510)814-2977
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 291 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-105-697-6

Query Match 38.1%; Score 40; DB 4; Length 291;
Best Local Similarity 61.5%; Pred. No. 62;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 7 FVGPDSRHSRH 19
Db 64 FVFDKAPSRH 76

RESULT 9
US-07-977-434-8
Sequence 8, Application US/07977434
Patent No. 5466591
GENERAL INFORMATION:
APPLICANT: Gelfand, David H.
APPLICANT: Abramson, Richard D.
TITLE OF INVENTION: 5' TO 3' EXONUCLEASE MUTATIONS OF
TITLE OF INVENTION: THERMOSTABLE DNA POLYMERASES
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: New Jersey
ZIP: 07110-1199
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: 7
SOFTWARE: Wordperfect 2.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/977,434
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 590,490
FILING DATE: 28-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 590,466
FILING DATE: 28-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 590,213

FILING DATE: 28-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 523,394
FILING DATE: 15-MAY-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 143,441
FILING DATE: 12-JAN-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 063,509
FILING DATE: 17-JUN-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 899,241
FILING DATE: 22-AUG-1986
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 746,121
FILING DATE: 15-AUG-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US90/07641
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 585,471
FILING DATE: 20-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 455,611
FILING DATE: 22-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 609,157
FILING DATE: 02-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 557,517
FILING DATE: 24-JUL-1990
ATTORNEY/AGENT INFORMATION:
NAME: Luann Cseri
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: Case No. 5466591 8753
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 814-2972
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 834 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-977-434-8

Query Match 38.1%; Score 40; DB 1; Length 834;
Best Local Similarity 61.5%; Pred. No. 2e+02;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 7 FVGPDSRHSRH 19
Db 64 FVFDKAPSRH 76

RESULT 10
US-07-977-434-10
Sequence 10, Application US/07977434
Patent No. 5466591
GENERAL INFORMATION:
APPLICANT: Gelfand, David H.
APPLICANT: Abramson, Richard D.
TITLE OF INVENTION: 5' TO 3' EXONUCLEASE MUTATIONS OF
TITLE OF INVENTION: THERMOSTABLE DNA POLYMERASES
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: New Jersey
ZIP: 07110-1199
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: Macintosh
OPERATING SYSTEM: 7
SOFTWARE: WordPerfect 2.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/977,434
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 590,490
FILING DATE: 28-SEP-1990
APPLICATION NUMBER: US 590,466
FILING DATE: 28-SEP-1990
APPLICATION NUMBER: US 590,213
FILING DATE: 28-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 523,394
FILING DATE: 15-MAY-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 143,441
FILING DATE: 12-JAN-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 063,509
FILING DATE: 17-JUN-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 899,241
FILING DATE: 22-AUG-1986
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 746,121
FILING DATE: 15-AUG-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US90/07641
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 585,471
FILING DATE: 20-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 455,611
FILING DATE: 22-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 609,157
FILING DATE: 02-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 557,517
FILING DATE: 24-JUL-1990
ATTORNEY/AGENT INFORMATION:
NAME: Luann Casert
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: Case No. 5466591 8753
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 814-2972
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 834 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-977-434-10

Query Match 38.1%; Score 40; DB 1; Length 834;
Best Local Similarity 61.5%; Pred. No. 2e+02;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 7 FVGFDASEPD5RH 19
| | | | |
Db 64 FVVFDAKAP5FRH 76

RESULT 11
US-08-073-384C-6
; Sequence 6, Application US/08073384C

Patent No. 5541311
GENERAL INFORMATION:
APPLICANT: Dahlberg, James E.
APPLICANT: Lyamichev, Victor I.
APPLICANT: Brow, Mary Ann D.
TITLE OF INVENTION: SYNTHESIS-DEFICIENT THERMOSTABLE DNA
TITLE OF INVENTION: POLYMERASE
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/073,384C
FILING DATE: 04-JUN-1993
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/986,330
FILING DATE: 07-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Peter G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: FORS-00613
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/705-8410
TELEFAX: 415/397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 834 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-073-384C-6

Query Match 38.1%; Score 40; DB 1; Length 834;
Best Local Similarity 61.5%; Pred. No. 2e+02;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 7 FVGFDASEPD5RH 19
| | | | |
Db 64 FVVFDAKAP5FRH 76

RESULT 12
US-08-254-359A-6
; Sequence 6, Application US/08254359A
Patent No. 5614402
GENERAL INFORMATION:
APPLICANT: DAHLBERG, JAMES E.
APPLICANT: LYAMICHEV, VICTOR I.
APPLICANT: BROW, MARY ANN D.
TITLE OF INVENTION: 5' NUCLEASES DERIVED FROM THERMOSTABLE
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/254,359A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/073,384
FILING DATE: 06-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/986,330
FILING DATE: 07-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: CARROLL, PETER G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: FORS-01000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 834 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-254-359A-6

Query Match 38.1%; Score 40; DB 1; Length 834;
Best Local Similarity 61.5%; Pred. No. 2e+02;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 7 FVGFDASEPDSSRH 19
DB 64 FVFDKAPSFRRH 76

RESULT 13
US-08-384-490-31
Sequence 31, Application US/08384490
Patent No. 5618711
GENERAL INFORMATION:
APPLICANT: Gelfand, David H.
APPLICANT: Lawyer, Frances C.
APPLICANT: Stofel, Susanne
TITLE OF INVENTION: Recombinant Expression Vectors and
TITLE OF INVENTION: Purification Methods for Thermophilus DNA
TITLE OF INVENTION: Polymerase
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/384,490
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/148,133
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Sias, Stacey R.
REGISTRATION NUMBER: 32,630

REFERENCE/DOCKET NUMBER: 8887
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 814-2863
TELEFAX: (510) 814-2977
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 834 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-384-490-31

Query Match 38.1%; Score 40; DB 1; Length 834;
Best Local Similarity 61.5%; Pred. No. 2e+02;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 7 FVGFDASEPDSSRH 19
DB 64 FVFDKAPSFRRH 76

RESULT 14
US-08-483-043-6
Sequence 6, Application US/08483043
Patent No. 5691142
GENERAL INFORMATION:
APPLICANT: Dahlberg, James E.
APPLICANT: Lyamichev, Victor I.
APPLICANT: Brow, Mary Ann D.
TITLE OF INVENTION: SYNTHESIS-DEFICIENT THERMOSTABLE DNA
TITLE OF INVENTION: POLYMERASE
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,043
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/073,384
FILING DATE: 04-JUN-1993
APPLICATION NUMBER: US 07/986,330
FILING DATE: 07-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Peter G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: FORS-00613
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/705-8410
TELEFAX: 415/397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 834 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-483-043-6

Query Match 38.1%; Score 40; DB 1; Length 834;

Best Local Similarity 61.5%; Pred. No. 2e+02;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 7 FVGFDASEPD SRH 19
DB 64 FVVFDAKAFSFRH 76

RESULT 15

US-08-459-383-31
Sequence 31, Application US/08459383
Patent No. 5741690
GENERAL INFORMATION:
APPLICANT: Gelfand, David H.
APPLICANT: Lawyer, Frances C.
APPLICANT: Stoffel, Susanne
TITLE OF INVENTION: Recombinant Expression Vectors and
TITLE OF INVENTION: Purification Methods for Thermophilus DNA
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,383
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/384,490
FILING DATE:
APPLICATION NUMBER: US/08/148,133
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Slas, Stacey R.
REGISTRATION NUMBER: 32,630
REFERENCE/DOCKET NUMBER: 8887
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 814-2863
TELEFAX: (510) 814-2977
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 834 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE:
US-08-459-383-31

Query Match 38.1%; Score 40; DB 1; Length 834;
Best Local Similarity 61.5%; Pred. No. 2e+02;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 7 FVGFDASEPD SRH 19
DB 64 FVVFDAKAFSFRH 76

Search completed: November 30, 2001, 09:43:38
Job time: 197 sec

Arabidopsis thaliana
Mustard raffinose
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Soybean raffinose
Soybean rhg1 pekín
Soybean rhg1 toyos
Soybean rhg1 noir
Soybean rhg1 lee a
Soybean rhg1 p100
Soybean rhg1 pekin
Soybean rhg1 toyos
Soybean rhg1 noir
Soybean rhg1 lee a
Soybean rhg1 p120
Soybean rhg1 pekín
Chicken ocysae rec
Peptide #6346 enco
Peptide #7535 enco
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Human secreted pro
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Streptococcus pneumoniae
Soybean 240017 reg

CC an optimum pH of 6 to 8 and working temperature of 35 to 40 degrees
 CC C, has a molecular weight of 75 to 95 kDa by gel filtration or 90
 CC to 100 kDa by PAGE and SDS-PAGE under reductive conditions and is
 CC inhibited by iodoacetamide, N-ethylmaleimide and myoinositol.
 CC
 XX
 SO Sequence 14 AA;

Query Match 100.0%; Score 80; DB 19; Length 14;
 Best Local Similarity 100.0%; Pred. No. 9.9e-08;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YDQDMVVVQVPWP 14
 |||
 DB 1 ydqdmvvvqvwp 14

RESULT 2

AA17424
 ID AAV17424 standard; peptide; 14 AA.

XX AC AAV17424;

DT 29-JUL-1999 (first entry)

DE Cucumber raffinose synthase peptide SEQ ID NO:3.

KW Raffinose synthase; sucrose; galactinol.

XX OS Cucumis sativus.

PN JP11123080-A.

XX PD 11-MAY-1999.

PF 24-OCT-1997; 97JP-0292969.

PR 24-OCT-1997; 97JP-0292969.

PA (AJIN) AJINOMOTO KK.

XX WPI; 1999-340516/29.

PT New raffinose synthase gene - for production of raffinose from
 sucrose and galactinol

PS Example 2; Page 22; 37pp; Japanese.

CC The present invention describes a raffinose synthase, having an activity
 CC of forming raffinose from sucrose and galactinol. The raffinose synthase
 CC gene can be used for expression in a plant for the production of
 CC raffinose. The raffinose synthase can give raffinose from sucrose
 CC and galactinol efficiently. The present sequence represents a raffinose
 CC synthase peptide from cucumber.
 CC
 XX

SO Sequence 14 AA;

Query Match 100.0%; Score 80; DB 20; Length 14;
 Best Local Similarity 100.0%; Pred. No. 9.9e-08;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YDQDMVVVQVPWP 14
 |||
 DB 1 ydqdmvvvqvwp 14

RESULT 3

AAW53570
 ID AAW53570 standard; Protein; 784 AA.

XX AC AAW53570;

XX

DT 06-JUL-1998 (first entry)

XX Cucumber raffinose synthase.

DE Cucumber; raffinose synthase; sucrose; galactinol.

XX OS Cucumis sativus.

PN JP10084973-A.

XX PD 07-APR-1998.

PF 28-APR-1997; 97JP-0111124.

PR 26-JUL-1996; 96JP-0198079.

XX PR 26-APR-1996; 96JP-0107682.

PA (AJIN) AJINOMOTO KK.

XX WPI; 1998-264858/24.

DR N-PSDB; AAV22250.

PT Raffinose synthase gene - useful for preparation of raffinose in
 transformed plant

PS Claim 3; Pages 17-20; 26pp; Japanese.

CC The present sequence is cucumber raffinose synthase, which
 CC forms raffinose from sucrose and galactinol, has an optimum pH of
 CC 6 to 8 and working temperature of 35 to 40 degrees C, has a
 CC molecular weight of 75 to 95 kDa by gel filtration or 90 to 100 kDa
 CC by PAGE and SDS-PAGE under reductive conditions and is inhibited by
 CC iodoacetamide, N-ethylmaleimide and myoinositol.
 CC
 XX

SO Sequence 784 AA;

Query Match 100.0%; Score 80; DB 19; Length 784;
 Best Local Similarity 100.0%; Pred. No. 8.8e-06;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YDQDMVVVQVPWP 14
 |||
 DB 756 ydqdmvvvqvwp 769

RESULT 4

AA17417
 ID AAV17417 standard; Protein; 784 AA.

XX AC AAV17417;

DT 29-JUL-1999 (first entry)

DE Cucumber raffinose synthase.

XX Raffinose synthase; sucrose; galactinol.

OS Cucumis sativus.

PN JP11123080-A.

XX PD 11-MAY-1999.

PF 24-OCT-1997; 97JP-0292969.

XX PR 24-OCT-1997; 97JP-0292969.

PA (AJIN) AJINOMOTO KK.

XX WPI; 1999-340516/29.

DR N-PSDB; AAX61238.


```
OS Glycine max.
XX
XX JP11215984-A.
XX
XX 10-AUG-1999.
XX
XX 12-DEC-1997; 97JP-0342899.
XX
XX 28-NOV-1997; 97JP-0329006.
XX
XX 18-DEC-1996; 96JP-0338673.
XX
XX (SUMO ) SUMITOMO CHEM CO LTD.
XX
XX WPI; 1999-51112/43.
XX
XX N-PSDB; AA210002.
XX
XX New raffinose synthase gene - is prepared from a plant material
XX
XX
XX Claim 8; Page 25-27; 40pp; Japanese.
XX
XX
XX The present sequence represents a raffinose synthase protein. The
XX
XX sequence is isolated from plant material. The protein forms raffinose
XX
XX by complexing alpha(1 to 6)- D-galactosyl hydroxyl group of the 6C
XX
XX of D-glucose residue in sucrose molecules.
XX
XX Sequence 781 AA;

OY 3 ODOMVVVQVWP 14
   :|:|: |||||
Db 758 edkmrvqvwp 769

RESULT 8
AAB49400
ID AAB49400 standard; Protein; 781 AA.
XX
XX AAB49400;
XX
XX 07-MAR-2001 (first entry)
XX
XX Soybean raffinose synthase.
XX
XX Plant promoter; transgenic plant; desired property.
XX
XX Glycine max.
XX
XX EP1048733-A2.
XX
XX 02-NOV-2000.
XX
XX 27-APR-2000; 2000EP-0108962.
XX
XX 30-APR-1999; 99JP-0124527.
XX
XX 01-SEP-1999; 99JP-0247211.
XX
XX (SUMO ) SUMITOMO CHEM CO LTD.
XX
XX Ishige F, Watanabe E, Oeda K;
XX
XX WPI; 2001-104537/12.
XX
XX N-PSDB; AAC89523.
XX
XX New soybean plant promoters useful for generating transgenic plants
XX
XX with desired properties -
XX
XX Example 6; Page 24-27; 36pp; English.
XX
XX The present invention provides novel plant promoters which can be used in
```

```
CC the production of transgenic plants which express genes with desired
XX
XX properties.
XX
XX Sequence 781 AA;

OY 3 ODOMVVVQVWP 14
   :|:|: |||||
Db 758 edkmrvqvwp 769

RESULT 9
AAG53157
ID AAG53157 standard; Protein; 484 AA.
XX
XX AAG53157;
XX
XX 18-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 67650.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
XX
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX
XX termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX
XX 05-MAR-1999; 99US-0123180.
XX
XX 09-MAR-1999; 99US-0123548.
XX
XX 23-MAR-1999; 99US-0125788.
XX
XX 25-MAR-1999; 99US-0126284.
XX
XX 29-MAR-1999; 99US-0126785.
XX
XX 01-APR-1999; 99US-0127462.
XX
XX 06-APR-1999; 99US-0128234.
XX
XX 08-APR-1999; 99US-0128714.
XX
XX 16-APR-1999; 99US-0129845.
XX
XX 19-APR-1999; 99US-0130077.
XX
XX 21-APR-1999; 99US-0130449.
XX
XX 23-APR-1999; 99US-0130510.
XX
XX 28-APR-1999; 99US-0130891.
XX
XX 30-APR-1999; 99US-0132048.
XX
XX 30-APR-1999; 99US-0132407.
XX
XX 04-MAY-1999; 99US-0132484.
XX
XX 05-MAY-1999; 99US-0132485.
XX
XX 06-MAY-1999; 99US-0132486.
XX
XX 07-MAY-1999; 99US-0132487.
XX
XX 11-MAY-1999; 99US-0132863.
XX
XX 14-MAY-1999; 99US-0134218.
XX
XX 14-MAY-1999; 99US-0134219.
XX
XX 14-MAY-1999; 99US-0134221.
XX
XX 14-MAY-1999; 99US-0134370.
XX
XX 18-MAY-1999; 99US-0134768.
XX
XX 19-MAY-1999; 99US-0134941.
XX
XX 20-MAY-1999; 99US-0135124.
XX
XX 21-MAY-1999; 99US-0135353.
XX
XX 24-MAY-1999; 99US-0135629.
XX
XX 25-MAY-1999; 99US-0136021.
XX
XX 27-MAY-1999; 99US-0136392.
XX
XX 28-MAY-1999; 99US-0136782.
XX
XX 01-JUN-1999; 99US-0137222.
```

PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145275.
PR 27-JUL-1999; 99US-0145912.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.

PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match

56.2%; Score 45; DB 21; Length 484;

Best Local Similarity 53.8%; Pred. No. 7.3;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
QY 2 DODMYVGVPM 14
|:|:|:|
Db 457 dedemivgddpwp 469

RESULT 10
AAG53156
ID AAG53156 standard; Protein: 563 AA.
XX AAG53156;
AC
XX 18-OCT-2000 (first entry)
DT XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 67649.
DE
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
OS
PN EPI033405-A2.
PD
XX 06-SEP-2000.
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 990S-0121825.
PR 05-MAR-1999; 990S-0123180.
PR 09-MAR-1999; 990S-0123548.
PR 23-MAR-1999; 990S-0125788.
PR 25-MAR-1999; 990S-0126264.
PR 29-MAR-1999; 990S-0126785.
PR 01-APR-1999; 990S-0127462.
PR 06-APR-1999; 990S-0128234.
PR 08-APR-1999; 990S-0128714.
PR 16-APR-1999; 990S-0129845.
PR 19-APR-1999; 990S-0130077.
PR 21-APR-1999; 990S-0130449.
PR 23-APR-1999; 990S-0130510.
PR 28-APR-1999; 990S-0130891.
PR 30-APR-1999; 990S-0131449.
PR 30-APR-1999; 990S-0132048.
PR 04-MAY-1999; 990S-0132407.
PR 05-MAY-1999; 990S-0132484.
PR 05-MAY-1999; 990S-0132485.
PR 06-MAY-1999; 990S-0132486.
PR 07-MAY-1999; 990S-0132487.
PR 11-MAY-1999; 990S-0134256.
PR 14-MAY-1999; 990S-0134218.
PR 14-MAY-1999; 990S-0134219.
PR 14-MAY-1999; 990S-0134221.
PR 18-MAY-1999; 990S-0134370.
PR 18-MAY-1999; 990S-0134768.
PR 19-MAY-1999; 990S-0134941.
PR 20-MAY-1999; 990S-0135124.
PR 21-MAY-1999; 990S-0135353.
PR 24-MAY-1999; 990S-0135629.
PR 25-MAY-1999; 990S-0136021.
PR 27-MAY-1999; 990S-0136392.
PR 28-MAY-1999; 990S-0136782.
PR 01-JUN-1999; 990S-0137222.
PR 03-JUN-1999; 990S-0137528.
PR 04-JUN-1999; 990S-0137502.
PR 07-JUN-1999; 990S-0137724.
PR 08-JUN-1999; 990S-0138094.
PR 10-JUN-1999; 990S-0138540.
PR 10-JUN-1999; 990S-0138847.
PR 14-JUN-1999; 990S-0139119.

PR 16-JUN-1999; 990S-0139452.
PR 16-JUN-1999; 990S-0139453.
PR 17-JUN-1999; 990S-0139492.
PR 18-JUN-1999; 990S-0139454.
PR 18-JUN-1999; 990S-0139455.
PR 18-JUN-1999; 990S-0139456.
PR 18-JUN-1999; 990S-0139457.
PR 18-JUN-1999; 990S-0139458.
PR 18-JUN-1999; 990S-0139459.
PR 18-JUN-1999; 990S-0139460.
PR 18-JUN-1999; 990S-0139461.
PR 18-JUN-1999; 990S-0139462.
PR 18-JUN-1999; 990S-0139463.
PR 18-JUN-1999; 990S-0139750.
PR 18-JUN-1999; 990S-0139763.
PR 21-JUN-1999; 990S-0139817.
PR 22-JUN-1999; 990S-0139899.
PR 23-JUN-1999; 990S-0140353.
PR 23-JUN-1999; 990S-0140354.
PR 24-JUN-1999; 990S-0140695.
PR 28-JUN-1999; 990S-0140823.
PR 29-JUN-1999; 990S-0140991.
PR 30-JUN-1999; 990S-0141287.
PR 01-JUL-1999; 990S-0141842.
PR 01-JUL-1999; 990S-0142154.
PR 02-JUL-1999; 990S-0142055.
PR 06-JUL-1999; 990S-0142390.
PR 08-JUL-1999; 990S-0142803.
PR 09-JUL-1999; 990S-0142920.
PR 12-JUL-1999; 990S-0142977.
PR 13-JUL-1999; 990S-0143542.
PR 14-JUL-1999; 990S-0143624.
PR 15-JUL-1999; 990S-0144005.
PR 16-JUL-1999; 990S-0144085.
PR 16-JUL-1999; 990S-0144086.
PR 19-JUL-1999; 990S-0144325.
PR 19-JUL-1999; 990S-0144331.
PR 19-JUL-1999; 990S-0144332.
PR 19-JUL-1999; 990S-0144333.
PR 19-JUL-1999; 990S-0144334.
PR 19-JUL-1999; 990S-0144335.
PR 20-JUL-1999; 990S-0144352.
PR 20-JUL-1999; 990S-0144684.
PR 21-JUL-1999; 990S-0144814.
PR 21-JUL-1999; 990S-0145086.
PR 21-JUL-1999; 990S-0145088.
PR 22-JUL-1999; 990S-0145085.
PR 22-JUL-1999; 990S-0145087.
PR 22-JUL-1999; 990S-0145089.
PR 22-JUL-1999; 990S-0145192.
PR 23-JUL-1999; 990S-0145145.
PR 23-JUL-1999; 990S-0145218.
PR 23-JUL-1999; 990S-0145224.
PR 26-JUL-1999; 990S-0145276.
PR 27-JUL-1999; 990S-0145913.
PR 27-JUL-1999; 990S-0145918.
PR 27-JUL-1999; 990S-0145919.
PR 28-JUL-1999; 990S-0145951.
PR 02-AUG-1999; 990S-0146386.
PR 02-AUG-1999; 990S-0146388.
PR 02-AUG-1999; 990S-0146389.
PR 03-AUG-1999; 990S-0147038.
PR 04-AUG-1999; 990S-0147204.
PR 04-AUG-1999; 990S-0147302.
PR 05-AUG-1999; 990S-0147192.
PR 05-AUG-1999; 990S-0147260.
PR 06-AUG-1999; 990S-0147303.
PR 06-AUG-1999; 990S-0147416.
PR 09-AUG-1999; 990S-0147493.
PR 09-AUG-1999; 990S-0147935.
PR 10-AUG-1999; 990S-0148171.
PR 11-AUG-1999; 990S-0148319.

Db 549 :|:|||||
eesmvmvqpw 559

RESULT 12

AAAG53155 standard; Protein; 586 AA.

AAAG53155;

18-OCT-2000 (first entry)

Arabidopsis thaliana protein fragment SEQ ID NO: 67648.

Protein identification: signal transduction pathway; metabolic pathway;
hybridisation assay: genetic mapping; gene expression control; promoter;
termination sequence.

Arabidopsis thaliana.

EP1033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-0301439.

25-FEB-1999; 99US-0121825.

05-MAR-1999; 99US-0123180.

09-MAR-1999; 99US-0123548.

22-MAR-1999; 99US-0125788.

23-MAR-1999; 99US-0126264.

29-MAR-1999; 99US-0126785.

01-APR-1999; 99US-0127462.

06-APR-1999; 99US-0128234.

08-APR-1999; 99US-0128714.

15-APR-1999; 99US-0129845.

19-APR-1999; 99US-0130077.

21-APR-1999; 99US-0130449.

23-APR-1999; 99US-0130510.

28-APR-1999; 99US-0130891.

30-APR-1999; 99US-0131449.

30-APR-1999; 99US-0132048.

04-MAY-1999; 99US-0132407.

05-MAY-1999; 99US-0132485.

06-MAY-1999; 99US-0132486.

07-MAY-1999; 99US-0132487.

11-MAY-1999; 99US-0132863.

14-MAY-1999; 99US-0134218.

14-MAY-1999; 99US-0134219.

14-MAY-1999; 99US-0134221.

18-MAY-1999; 99US-0134768.

19-MAY-1999; 99US-0134941.

20-MAY-1999; 99US-0135124.

21-MAY-1999; 99US-0135533.

24-MAY-1999; 99US-0135629.

25-MAY-1999; 99US-0136021.

27-MAY-1999; 99US-0136392.

28-MAY-1999; 99US-0136782.

01-JUN-1999; 99US-0137222.

03-JUN-1999; 99US-0137528.

04-JUN-1999; 99US-0137502.

07-JUN-1999; 99US-0137724.

08-JUN-1999; 99US-0138094.

10-JUN-1999; 99US-0138540.

10-JUN-1999; 99US-0138847.

14-JUN-1999; 99US-0139119.

16-JUN-1999; 99US-0139452.

16-JUN-1999; 99US-0139453.

17-JUN-1999; 99US-0139492.

18-JUN-1999; 99US-0139454.

18-JUN-1999; 99US-0139455.

18-JUN-1999; 99US-0139456.

18-JUN-1999; 99US-0139457.

18-JUN-1999; 99US-0139458.

18-JUN-1999; 99US-0139459.

18-JUN-1999; 99US-0139460.

18-JUN-1999; 99US-0139461.

18-JUN-1999; 99US-0139462.

18-JUN-1999; 99US-0139463.

18-JUN-1999; 99US-0139750.

18-JUN-1999; 99US-0139763.

21-JUN-1999; 99US-0139817.

22-JUN-1999; 99US-0139899.

23-JUN-1999; 99US-0140353.

23-JUN-1999; 99US-0140354.

24-JUN-1999; 99US-0140695.

28-JUN-1999; 99US-0140823.

29-JUN-1999; 99US-0140981.

30-JUN-1999; 99US-0141287.

01-JUL-1999; 99US-0141842.

01-JUL-1999; 99US-0142154.

02-JUL-1999; 99US-0142055.

06-JUL-1999; 99US-0142390.

08-JUL-1999; 99US-0142803.

09-JUL-1999; 99US-0142820.

12-JUL-1999; 99US-0142977.

13-JUL-1999; 99US-0143542.

14-JUL-1999; 99US-0143624.

15-JUL-1999; 99US-0144005.

16-JUL-1999; 99US-0144085.

16-JUL-1999; 99US-0144825.

19-JUL-1999; 99US-0144331.

19-JUL-1999; 99US-0144332.

19-JUL-1999; 99US-0144333.

19-JUL-1999; 99US-0144334.

20-JUL-1999; 99US-0144335.

20-JUL-1999; 99US-0144352.

20-JUL-1999; 99US-0144632.

21-JUL-1999; 99US-0144884.

21-JUL-1999; 99US-0144814.

21-JUL-1999; 99US-0145086.

21-JUL-1999; 99US-0145088.

22-JUL-1999; 99US-0145085.

22-JUL-1999; 99US-0145087.

22-JUL-1999; 99US-0145089.

22-JUL-1999; 99US-0145192.

23-JUL-1999; 99US-0145145.

23-JUL-1999; 99US-0145218.

23-JUL-1999; 99US-0145224.

26-JUL-1999; 99US-0145276.

27-JUL-1999; 99US-0145513.

27-JUL-1999; 99US-0145918.

27-JUL-1999; 99US-0145919.

28-JUL-1999; 99US-0145951.

02-AUG-1999; 99US-0146386.

02-AUG-1999; 99US-0146388.

02-AUG-1999; 99US-0146389.

03-AUG-1999; 99US-0147038.

04-AUG-1999; 99US-0147204.

04-AUG-1999; 99US-0147302.

05-AUG-1999; 99US-0147192.

05-AUG-1999; 99US-0147260.

06-AUG-1999; 99US-0147303.

06-AUG-1999; 99US-0147416.

09-AUG-1999; 99US-0147493.

09-AUG-1999; 99US-0147935.

10-AUG-1999; 99US-0148171.

11-AUG-1999; 99US-0148319.

12-AUG-1999; 99US-0148341.

13-AUG-1999; 99US-0148565.

13-AUG-1999; 99US-0148564.

16-AUG-1999; 99US-0149368.


```
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

```
Query Match 56.2%; Score 45; DB 21; Length 586;
Best Local Similarity 53.8%; Pred. No. 9.1;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
```

```
OY 2 DODMNVVQVWP 14
DB 559 dedemlvqddpwp 571
```

```
RESULT 13
AAV32074
ID AAV32074 standard; Protein; 777 AA.
```

```
XX
AC AAV32074;
XX
DT 17-JAN-2000 (first entry)
XX
DE Mustard raffinose synthase.
XX
KW Raffinose synthase; mustard; transgenic plant.
XX
OS Brassica juncea.
XX
FH Key Location/Qualifiers
FT Misc-difference 210
FT /note= "encoded by ACR"
XX
PN EP953643-A2.
XX
PD 03-NOV-1999.
XX
PF 27-APR-1999; 99EP-0107430.
XX
PR 30-APR-1998; 98JP-0120550.
PR 30-APR-1998; 98JP-0120551.
PR 04-DEC-1998; 98JP-0345590.
PR 10-DEC-1998; 98JP-0351246.
XX
PA (SUMO ) SUMITOMO CHEM CO LTD.
XX
PI Watanabe E, Oeda K;
XX
PI MPI: 1999-593144/51.
XX
DR N-PSDB; AA220209.
XX
PT New sense and antisense genes, useful for altering the level of
PT raffinose in food plants -
XX
PS Claim 26; Page 29-31; 55pp; English.
XX
CC This sequence represents mustard raffinose synthase, a protein
CC that can bind a D-galactosyl group through an alpha(1-6) bond to the
CC hydroxy group attached to the carbon atom at the 6-position of the
CC D-glucose residue in a sucrose molecule to form raffinose. cDNA
CC (see AA220209) encoding the enzyme was isolated from mustard
CC (Brassica juncea) leaf cDNA by PCR. Probes or primers generated from
CC plant raffinose synthase genes (see AA220207-10) may be used to obtain
CC other raffinose synthase genes by labeled detection or amplification
CC (claimed). These genes may be used to control the levels of
CC raffinose produced in plants. Antisense genes can be used to knock
CC out existing gene activity, and sense genes to increase the level
CC of gene activity. The resulting transgenic plants may be used as a
CC food source to alter the growing conditions for gut enterobacteria,
CC providing general health advantages.
XX
SQ Sequence 777 AA;
```

```
Query Match 56.2%; Score 45; DB 20; Length 777;
Best Local Similarity 63.6%; Pred. No. 12;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
```

```
OY 3 QDOMNVVQVWP 13
DB 754 eesmvmvqvpw 764
```

```
RESULT 14
AAG07534
ID AAG07534 standard; Protein; 228 AA.
XX
AC AAG07534;
XX
DT 17-OCT-2000 (first entry)
XX
```

DE Arabidopsis thaliana protein fragment SEQ ID NO: 4725.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX
PD 06-SEP-2000.
XX
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 30-APR-1999; 99US-0132484.
PR 04-MAY-1999; 99US-0132485.
PR 05-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139482.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 19-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 22-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147260.
PR 05-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 06-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 09-AUG-1999; 99US-0148171.
PR 10-AUG-1999; 99US-0148319.
PR 11-AUG-1999; 99US-0148341.
PR 12-AUG-1999; 99US-0148365.
PR 13-AUG-1999; 99US-0148684.
PR 13-AUG-1999; 99US-0149368.
PR 16-AUG-1999; 99US-0149175.
PR 17-AUG-1999; 99US-0149426.
PR 18-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.

PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0158293.
 PR 13-OCT-1999; 99US-0158294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 55.0%; Score 44; DB 21; Length 228;
 Best Local Similarity 61.5%; Pred. No. 4.8;
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 DDDQWVGVVWVMP 14
 DB 184 dgdmlvgdvmp 196

RESULT 15
 AAG45847
 ID AAG45847 standard; Protein: 228 AA.
 AC AAG45847;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 57609.

KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.

OS Arabidopsis thaliana.

XX EPI033405-A2.
 PN
 XX 06-SEP-2000.
 PD
 XX
 PF 25-FEB-2000; 2000EP-0301439.
 XX
 PR 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 23-APR-1999; 99US-0130891.
 PR 28-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 06-MAY-1999; 99US-0132487.
 PR 07-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.

PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 55.0%; Score 44; DB 21; Length 228;
Best Local Similarity 61.5%; Pred. No. 4.8;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
OY 2 DODMVTVOVPWP 14
| | | | |
Db 184 dgdmwlvgtvvpwp 196

Search completed: November 30, 2001, 09:42:57
Job time: 156 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 30, 2001, 09:41:42 ; Search time 39.46 seconds
(without alignments)
27.026 Million cell updates/sec

Title: US-08-846-234-3

Perfect score: 80

Sequence: 1 YDQDQWVGVVWP 14

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44	55.0	164	2	S58501 auxin-induced prot
2	44	55.0	225	2	H71407 auxin-induced prot
3	44	55.0	229	2	H86173 hypothetical prot
4	44	55.0	234	2	C85159 IAA7 like protein
5	44	55.0	619	2	C86467 hypothetical prot
6	44	55.0	719	2	T47727 hypothetical prot
7	43	53.8	423	2	T52251 5'-adenylylsulfate
8	43	53.8	863	1	S51789 VLDL receptor prec
9	42	52.5	500	2	D86274 hypothetical prote
10	42	52.5	529	2	T42584 tegument protein 4
11	42	52.5	600	2	C86468 probable auxin res
12	42	52.5	767	2	G86476 protein F1504.37 (
13	41	51.2	530	1	W2BED4 gene 40 protein -
14	40	50.0	189	2	S58493 auxin-induced prot
15	40	50.0	189	2	E36569 probable IAA6 prot
16	40	50.0	653	2	S75763 acetate--CoA ligase
17	40	50.0	1159	2	T43461 probable phosphol
18	39	48.8	206	2	S22400 insecticynanin A -
19	39	48.8	281	2	A65219 phnD protein - Bsc
20	39	48.8	281	2	F86104 phosphonate metad
21	39	48.8	305	2	T08963 phytochrome-assoc
22	39	48.8	384	2	G82670 general secretory
23	39	48.8	848	2	JG0194 androgen receptor
24	38	47.5	162	2	C82765 conserved hypothet
25	38	47.5	190	2	E64950 yeast protein - Bsc
26	38	47.5	190	2	E65800 hypothetical prote
27	38	47.5	305	2	A25010 pectinesterase (EC
28	38	47.5	390	2	A70656 hypothetical prote
29	38	47.5	432	2	S22527 glutamate--ammonia

30	38	47.5	473	2	E84471 probable beta-1,3-
31	38	47.5	524	2	D82220 conserved hypothet
32	38	47.5	624	2	B96556 hypothetical prote
33	38	47.5	722	2	T30995 hypothetical prote
34	38	47.5	1206	2	S72620 probable reverse t
35	38	47.5	1663	2	T42092 s-afadin - rat
36	38	47.5	1687	2	T30244 phosphodiesterase
37	38	47.5	1706	2	T30175 exoribonuclease, v
38	38	47.5	1719	2	T30174 exoribonuclease, v
39	38	47.5	1829	2	T41751 l-afadin - rat
40	37	46.2	166	2	S35795 androgen receptor
41	37	46.2	168	2	A71408 auxin-induced prot
42	37	46.2	179	2	S39078 auxin-induced prot
43	37	46.2	179	2	F96835 unknown protein F5
44	37	46.2	188	2	T10859 auxin-induced prot
45	37	46.2	189	2	S39075 auxin-induced prot

ALIGNMENTS

RESULT 1
S58501
auxin-induced protein IAA14 - Arabidopsis thaliana (fragment)
N:Alternate names: indoleacetic acid-inducible protein IAA14
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 22-Nov-1996 #sequence_revision 22-Nov-1996 #text_change 20-Aug-1999
C:Accession: S58501; S71408
R:Abel, S.; Nguyen, M.D.; Theologis, A.
J. Mol. Biol. 251, 533-549, 1995
A>Title: The PS-IAA4/5-like family of early auxin-inducible mRNAs in Arabidopsis thal
A:Reference number: S58491; MUID:95587993
A:Accession: S58501
A>Status: nucleic acid sequence not shown
A:Molecule type: mRNA
R:Theologis, A.
A:Residues: 1-164 <ABE>
submitted to the EMBL Data Library, December 1994
A:Reference number: S71406
A:Accession: S71408
A:Molecule type: DNA
A:Residues: 1-109, 'S', 111-164 <THE>
A:Cross-references: EMBL:U18416; NID:g972930; PIDN:AAC49055.1; PID:g972931
C:Genetics:
A:Gene: IAA14
C:Superfamily: auxin-induced protein aux28

Query Match 55.0% Score 44: DB 2: Length 164;
Best Local Similarity 61.5% Pred No. 1.6;
Matches 8: Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 DQDQWVGVVWP 14
DB 120 DQDQWVGVVWP 132

RESULT 2
H71407
auxin-induced protein - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
A:Variety: Columbia
C>Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 20-Jun-2000
C:Accession: H71407
R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; D
P.; Medler, H.; Wedler, E.; Wambutt, R.; Wellenreger, T.; Pohl, T.M.; Terry, N.; G
avanagh, T.; Hempel, S.; Kotter, P.; Ertlan, K.D.; Rieger, M.; Schaeffer, M.; Funk, B
Nature 391, 485-488, 1998
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Pulgaden
ethoff, A.; Moeres, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.;
C.; Chalvatzis, N.
A>Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis t
A:Reference number: A71400; MUID:96121113

A:Accession: H71407
A:Status: Preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-225 <REV>
A:Cross-references: GB:297336; NID:g2244788; PIDN:CABI0234.1; PID:g2244811
C:Genetics:
A:Map position: 4COP9-4G3845
C:Superfamily: auxin-induced protein aux28

Query Match 55.0%; Score 44; DB 2; Length 225;
Best Local Similarity 61.5%; Pred. No. 2.2;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 2 DODQMVVQVWP 14
| | | | |
Db 184 DGDMLVGDVWP 196

RESULT 3
H86173
hypothetical protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: H86173
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, R.; Theologis, A.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Jensen, N.F.; Hughes, B.; Hultzer, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: H86173
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-229 <STO>
A:Cross-references: GB:AE005172; NID:g4389514; PIDN:AAB70451.2; GSPDB:GN00141
C:Genetics:
A:Map position: 1
C:Superfamily: auxin-induced protein aux28

Query Match 55.0%; Score 44; DB 2; Length 229;
Best Local Similarity 61.5%; Pred. No. 2.3;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 2 DODQMVVQVWP 14
| | | | |
Db 185 DGDMLVGDVWP 197

RESULT 4
C85159
IAA7 like protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001
C:Accession: C85159
R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring Nature 402, 769-777, 1999
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A:Reference number: A85001; MUID:20083488
A:Accession: C85159
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-234 <STO>
A:Cross-references: GB:NC_001268; NID:g5302771; PIDN:CAB46059.1; GSPDB:GN00140
C:Genetics:
A:Gene: d13315c
A:Map position: 4

Query Match 55.0%; Score 44; DB 2; Length 234;
Best Local Similarity 61.5%; Pred. No. 2.3;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 2 DODQMVVQVWP 14
| | | | |
Db 184 DGDMLVGDVWP 196

RESULT 5
H86467
hypothetical protein AAD39615.1 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: C86467
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, R.; Theologis, A.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, Jensen, N.F.; Hughes, B.; Hultzer, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: C86467
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-619 <STO>
A:Cross-references: GB:AE005172; NID:g5091627; PIDN:AAD39615.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match 55.0%; Score 44; DB 2; Length 619;
Best Local Similarity 53.8%; Pred. No. 7;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 2 DODQMVVQVWP 14
| | | | |
Db 592 DEDKMLVGDVWP 604

RESULT 6
T47727
hypothetical protein F18021.60 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C>Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 02-Sep-2000
C:Accession: T47727
R:Benes, V.; Wurnbach, E.; Drzonek, H.; Ansoerge, W.; Mewes, H.W.; Rudd, S.; Lemcke, K submitted to the Protein Sequence Database, April 2000
A:Reference number: Z24474
A:Accession: T47727
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-719 <BEN>
A:Cross-references: EMBL:AF18021
A:Experimental source: cultivar Columbia; BAC clone F18021
C:Genetics:
A:Map position: 3
A:introns: 244/2; 491/1; 511/1
A:Note: F18021.60
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom

Query Match 55.0%; Score 44; DB 2; Length 719;
Best Local Similarity 54.5%; Pred. No. 8.2;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 3 QDQMVVQVWP 13

1 : : : : :
Db 100 OCGVIVQLPW 110
RESULT 7
T52251
5'-adenylylsulfate reductase (EC 1.8.99.-) [validated] - green alga (Enteromorpha intest
C:Species: Enteromorpha intestinalis (hollow green seaweed)
C.Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 20-Oct-2000
C:Accession: T52251
R:Geo, Y.; Schofield, O.M.; Leustek, T.
Plant Physiol. 123, 1087-1096, 2000
A:Title: Characterization of sulfate assimilation in marine algae focusing on the enzyme
A:Reference number: 226003; MUID:20349720
A:Accession: T52251
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-423 <GAO>
A:Cross-references: EMBL:AF069951; PIDN:AAC26855.1
C:Function:
A:Description: catalyzes reduced glutathione-dependent reduction of adenylylsulfate to S
C:Keywords: oxidoreductase
Query Match 53.8%; Score 43; DB 2; Length 423;
Best Local Similarity 58.3%; Pred. No. 6.9;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
Oy 2 DDDQMYVGVWP 13
1 : : : : :
Db 330 DKDTMVLVAPW 341
RESULT 8
S51789
VLDL receptor precursor - chicken
N:Alternate names: very low density lipoprotein receptor; vitellogenin receptor
C:Species: Gallus gallus (chicken)
C.Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 16-Jul-1999
C:Accession: S51789
R:Bujo, H.; Hermann, M.; Kaderli, M.O.; Jacobsen, L.; Sugawara, S.; Nimpf, J.; Yamamoto,
EMBO J. 13, 5165-5175, 1994
A:Title: Chicken oocyte growth is mediated by an eight ligand binding repeat member of t
A:Reference number: S51789; MUID:95045409
A:Accession: S51789
A:Molecule type: mRNA
A:Residues: 1-863 <BUJ>
A:Cross-references: EMBL:X80207; NID:g609265; PIDN:CA56505.1; PID:g609266
C:Comment: This receptor mediates uptake of very low density lipoprotein and vitellogenin
C:Comment: There is some evidence for the existence of a longer splice form containing a
C:Superfamily: LDL receptor; EGF homology; LDL receptor ligand-binding repeat homology;
C:Keywords: duplication; fatty acid metabolism; glycoprotein; receptor; transmembrane pr
F:1-44/Domain: signal sequence status predicted <SIG>
F:45-863/Product: VLDL receptor status predicted <MUT>
F:45-87/Domain: extracellular status predicted <EXT>
F:51-85/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:90-126/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F:131-167/Domain: LDL receptor ligand-binding repeat homology <LDL3>
F:172-206/Domain: LDL receptor ligand-binding repeat homology <LDL4>
F:211-247/Domain: LDL receptor ligand-binding repeat homology <LDL5>
F:257-291/Domain: LDL receptor ligand-binding repeat homology <LDL6>
F:296-330/Domain: LDL receptor ligand-binding repeat homology <LDL7>
F:336-373/Domain: LDL receptor ligand-binding repeat homology <LDL8>
F:378-412/Domain: EGF homology <EG1>
F:418-452/Domain: EGF homology <EG2>
F:459-498/Domain: LDL receptor WTD-containing repeat homology <Y1>
F:499-544/Domain: LDL receptor WTD-containing repeat homology <Y2>
F:545-587/Domain: LDL receptor WTD-containing repeat homology <Y3>
F:588-631/Domain: LDL receptor WTD-containing repeat homology <Y4>
F:632-674/Domain: LDL receptor WTD-containing repeat homology <Y5>
F:675-711/Domain: LDL receptor WTD-containing repeat homology <Y6>
F:726-769/Domain: EGF homology <EG3>
F:788-809/Domain: transmembrane status predicted <TM>

F:824-828/Region: coated-pit mediated internalization signal
F:169, 773/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:378-389,385-398,400-412,418-428,424-437,439-452,726-739,735-754,756-769/DLSulfide b
Query Match 53.8%; Score 43; DB 1; Length 863;
Best Local Similarity 35.7%; Pred. No. 15;
Matches 5; Conservative 5; Mismatches 4; Indels 0; Gaps 0;
Oy 1 YDDQMYVGVWP 14
1 : : : : :
Db 609 FDRQLVTEICWP 622
RESULT 9
D86274
hypothetical protein AAD39294.1 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C.Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: D86274
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Hughes, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: D86274
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-500 <STO>
A:Cross-references: GB:AE005172; NID:g5080784; PIDN:AAD39294.1; GSPDB:GN00141
A:Genetics:
A:Map position: 1
Query Match 52.5%; Score 42; DB 2; Length 500;
Best Local Similarity 50.0%; Pred. No. 12;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
Oy 2 DDDQMYVGVWP 13
1 : : : : :
Db 225 ESPQILRIQWP 236
RESULT 10
T42584
tegument protein 40 - equine herpesvirus 4 (strain NS80567)
C:Species: equine herpesvirus 4
A:Variety: strain NS80567
C.Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000
C:Accession: T42584
R:Reilford, E.A.; Watson, M.S.; Perry, J.; Cullinane, A.A.; Davison, A.J.
J. Gen. Virol. 79, 1197-1203, 1998
A:Title: The DNA sequence of equine herpesvirus-4.
A:Reference number: 222173; MUID:98264497
A:Accession: T42584
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-529 <TEL>
A:Cross-references: EMBL:AF030027; NID:g2605950; PIDN:AAC59558.1; PID:g2605986
A:Experimental source: strain NS80567
A:Genetics:
A:Gene: 40
C:Superfamily: varicella-zoster virus gene 38 protein
Query Match 52.5%; Score 42; DB 2; Length 529;
Best Local Similarity 45.5%; Pred. No. 13;

Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0; *

QY 4 DQDMVVVQVWP 14
: : : : :
Db 283 DEILVEEPPW 293

RESULT 11
C86468
Probable auxin response factor, 53188-50111 [Imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: C86468
R:theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Hultzer, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
A.; Li, J.H.; Li, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maitl, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: C86468
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-600 <STO>
A:Cross-references: GB:AE005172; NID:g10092376; PIDN:AA612783.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match 52.5%; Score 42; DB 2; Length 600;
Best Local Similarity 53.8%; Pred. No. 15;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 DQDMVVVQVWP 14
: : : : :
Db 566 DDKMLVGDPPW 578

RESULT 12
G86476
protein F1504.37 [Imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: G86476
R:theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Hultzer, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
A.; Li, J.H.; Li, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maitl, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: G86476
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-767 <STO>
A:Cross-references: GB:AE005172; NID:g8778363; PIDN:AAF9371.1; GSPDB:GN00141
C:Genetics:
A:Gene: F1504.37
A:Map position: 1

Query Match 52.5%; Score 42; DB 2; Length 767;
Best Local Similarity 46.2%; Pred. No. 20;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 DQDMVVVQVWP 14
: : : : :
Db 621 EEDKMLVGDPPW 633

RESULT 13
WZBED4
gene 40 protein - equine herpesvirus 1 (strain Ab4p)
C:Species: equine herpesvirus 1
A:Note: host Equus caballus (domestic horse)
C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 16-Jul-1999
C:Accession: F36799
R:Telford, E.A.R.; Watson, M.S.; McBride, K.; Davison, A.J.
submitted to Genbank, March 1992
A:Description: The DNA sequence of equine herpesvirus-1.
A:Reference number: A36805
A:Accession: F36799
A:Molecule type: DNA
A:Residues: 1-530 <TEL>
A:Cross-references: GB:M86664; NID:g330791; PIDN:AA602476.1; PID:g330833
R:Telford, E.A.R.; Watson, M.S.; McBride, K.; Davison, A.J.
Virology 189, 304-316, 1992
A:Title: The DNA sequence of equine herpesvirus-1.
A:Reference number: A41831; MUID:9229556
A:Contents: annotation; possible protein-coding frames
A:Note: neither amino acid nor nucleotide sequence is given
C:Genetics:
A:Gene: 40
C:Superfamily: varicella-zoster virus gene 38 protein

Query Match 51.2%; Score 41; DB 1; Length 530;
Best Local Similarity 36.4%; Pred. No. 20;
Matches 4; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY 4 DQMVVVQVWP 14
: : : : :
Db 284 DEILVEEPPW 294

RESULT 14
S58493
auxin-induced protein IAA6 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 29-Nov-1995 #sequence_revision 01-Mar-1996 #text_change 20-Aug-1999
C:Accession: S58493
R:Abel, S.; Nguyen, M.D.; Theologis, A.
J. Mol. Biol. 251, 533-549, 1995
A:Title: The PS-IAA4/5-like family of early auxin-inducible mRNAs in Arabidopsis thal
A:Reference number: S58491; MUID:95387393
A:Accession: S58493
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-189 <ABE>
A:Cross-references: EMBL:U018408; NID:g972914; PIDN:AA649047.1; PID:g972915
C:Genetics:
A:Gene: IAA6
C:Superfamily: auxin-induced protein aux28
C:Keywords: nucleus

Query Match 50.0%; Score 40; DB 2; Length 189;
Best Local Similarity 58.3%; Pred. No. 9.5;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 DQDMVVVQVWP 13
: : : : :
Db 152 DRDMMLVGDPPW 163

RESULT 15
E96569
probable IAA6 protein, 42631-41742 [Imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
 C:Accession: E96569
 R:Theologis, A.; Becker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
 ansen, N.F.; Hughes, B.; Hulzar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luos, J.S.; Malt, R.; Marziani,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719
 A:Accession: E96569
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-189 <STO>
 A:Cross-references: GB:AE005173; NID:g6682944; PIDN:AAF30332.1; GSPDB:GN00141
 C:Genetics:
 A:Gene: FLAG24.10
 A:Map position: 1
 C:Superfamily: auxin-induced protein aux28

Query Match 50.0%; Score 40; DB 2; Length 189;
 Best Local Similarity 58.3%; Pred. No. 9.5;
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 2 DQDQWVVGVPW 13
 1:|:|:|
 Db 152 DRDWMVGVDPW 163

Search completed: November 30, 2001, 09:41:43
 Job time: 82 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 30, 2001, 09:40:55 ; Search time 23.6 Seconds
(without alignments)
21.750 Million cell updates/sec

Title: US-08-846-234-3
Perfect score: 80
Sequence: 1 YDQDMVYGVMP 14

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SWISSPROT_39:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44	55.0	229	1	AX1H_ARATH
2	43	53.8	863	1	LDVR_CHICK
3	41	51.2	530	1	UL21_HSVB
4	40	50.0	189	1	AX16_ARATH
5	40	50.0	653	1	ACSA_SYNY3
6	39	48.8	281	1	PHMJ_ECOLI
7	38	47.5	188	1	YECW_ECOLI
8	38	47.5	528	1	DYR2_HUMAN
9	38	47.5	574	1	ERGL_HUMAN
10	37	46.2	168	1	AX11_ARATH
11	37	46.2	174	1	AX12_ARATH
12	37	46.2	179	1	AX16_PEA
13	37	46.2	186	1	AX14_ARATH
14	37	46.2	188	1	AX2C_PHAU
15	37	46.2	189	1	AX14_PEA
16	37	46.2	193	1	AX2D_PHAU
17	37	46.2	194	1	AX2A_PHAU
18	37	46.2	195	1	AX22_SOYB
19	37	46.2	196	1	AX2B_PHAU
20	37	46.2	203	1	AX2E_PHAU
21	37	46.2	236	1	AX1G_ARATH
22	37	46.2	243	1	AX28_SOYB
23	37	46.2	243	1	AX17_ARATH
24	37	46.2	251	1	SNST_AMAPL
25	37	46.2	261	1	AX1A_ARATH
26	37	46.2	321	1	AX18_ARATH
27	37	46.2	338	1	AX19_ARATH
28	37	46.2	373	1	GSPL_XANCP
29	37	46.2	428	1	GLN2_ORYSA
30	37	46.2	475	1	DXR_MENPT
31	37	46.2	477	1	DXR_HUMAN
32	37	46.2	532	1	GRB7_HUMAN
33	37	46.2	709	1	ANDR_RABIT

ALIGNMENTS

RESULT	ID	AX1H_ARATH	STANDARD	PRT	229 AA
AC	P93830				
DT	20-AUG-2001	(Rel. 40, Created)			
DT	20-AUG-2001	(Rel. 40, Last sequence update)			
DE	AUXIN-RESPONSIVE PROTEIN IAA17 (INDOLEACETIC ACID-INDUCED PROTEIN 17).				
GN	IAA17 OR AXR3 OR F19P19.31.				
OS	Arabidopsis thaliana (Mouse-ear cress).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;				
OC	eucosids II; Brassicales; Brassicaceae; Arabidopsi.				
OX	NCBI_TaxID=3702;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=CV. COLUMBIA;				
RX	MEDLINE=98004476; PubMed=9342315;				
RA	Kim J., Harter K., Theologis A.;				
RT	*Protein-protein interactions among the Aux/IAA proteins.*;				
RL	Proc. Natl. Acad. Sci. U.S.A. 94:11786-11791(1997).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=CV. COLUMBIA;				
RX	MEDLINE=98146427; PubMed=9478901;				
RA	Rouse D., Mackay P., Stirnberg P., Estelle M., Leyser O.;				
RT	*Changes in auxin response from mutations in an Aux/IAA gene.*;				
RL	Science 279:1371-1373(1998).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=CV. COLUMBIA;				
RA	Vysotskaia V.S., Osborne B.I., Toriumi M., Yu G., Oji O., Buehler E.,				
RA	Conway A.B., Conway A.R., Dewar K., Feng J., Kim C., Kurtz D., Li Y.,				
RA	Shinn P., Sun H., Davis R.W., Ecker J.R., Federspiel N.A.,				
RA	Theologis A.;				
RT	Submitted (Jan-1997) to the EMBL/GenBank/DBJ databases.				
CC	-1- FUNCTION: COULD ACT AS REGULATOR OF GENES RESPONSIBLE FOR				
CC	MEDIATING THE VARIOUS AUXIN-INDUCED EVENTS RESPONSIBLE FOR CELL				
CC	GROWTH (BY SIMILARITY).				
CC	-1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).				
CC	-1- INDUCTION: BY AUXIN.				
CC	-1- SIMILARITY: BELONGS TO THE AUX/IAA FAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sdb.ch/announce/				
CC	or send an email to license@sdb-sib.ch).				
CC	-----				
DR	EMBL: U49073; AB84354.1; -				P35953 oryctolagus
DR	EMBL: AC000104; AAB70451.2; -				O97776 eulemur ful
DR	EMBL: AF040631; AAC39439.1; -				O97952 macaca fasc
DR	Mendel: 7119; Arath:1524;7119.				O97960 papio hamad
DR	Interpro: IPR003311; AUX_1AA.				P15091 mus musculu
					P15207 rattus norv
					O9t90 canis famill
					O97775 pan troglod
					P10275 homo sapien
					P20043 lactobacill
					O9vm75 drosophila
					O9eps2 mus musculu

Query Match 53.8%; Score 43; DB 1; Length 863;
 RT Best Local Similarity 35.7%; Pred. No. 6.4;
 Matches 5; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

OY 1 YDQDQVVVQVPM 14
 ID 609 FDRQQLVTEIQMP 622

RESULT 3
 ID1_HSVB STANDARD; PRT; 530 AA.
 AC P28972;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 01-DEC-1992 (Rel. 24, Last annotation update)
 DE GENE 40 PROTEIN.
 GN 40.
 OS Equine herpesvirus type 1 (strain Abap) (EHV-1).
 OC Viruses; deDNA viruses, no RNA stage; Herpesviridae;
 OC Alphaherpesvirinae; Varicelloviruses.
 OX NCBI_TaxID=31520;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92295566; PubMed=1318606;
 RA Telford E.A.R., Watson M.S., McBride K., Davison A.J.;
 RT "The DNA sequence of equine herpesvirus-1";
 RL Virology 189:304-316(1992).
 CC -1 SIMILARITY: BELONGS TO FAMILY THAP GROUPS TOGETHER HSV-1 UL21,
 EHV-1 40, EHV-4 UL21, AND VZV 38.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; M86664; AAB02476.1; -
 DR PIR; F36799; W2BED4.
 SQ SEQUENCE 530 AA; 57915 MW; 2E43565AB297CD18 CRC64;

Query Match 51.2%; Score 41; DB 1; Length 530;
 RT Best Local Similarity 36.4%; Pred. No. 8.6;
 Matches 4; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

OY 4 DQVVVQVPM 14
 ID 284 DETILIEPMP 294

RESULT 4
 ID AX16_ARATH STANDARD; PRT; 189 AA.
 AC Q38824;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE AUXIN-RESPONSIVE PROTEIN IAA6 (INDOLEACETIC ACID-INDUCED PROTEIN 6).
 GN IAA6
 OS Arabidopsis thaliana (Mouse-ear cress).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=95387393; PubMed=7658471;
 RA Abel S., Nguyen M.D., Theologis A.;

RT "The PS-IAA4/5-like family of early auxin-inducible mRNAs in
 RT Arabidopsis thaliana."
 RL J. Mol. Biol. 251:533-549(1995).
 CC -1- FUNCTION: COULD ACT AS REGULATOR OF GENES RESPONSIBLE FOR CELL
 CC MEDIATING THE VARIOUS AUXIN-INDUCED EVENTS RESPONSIBLE FOR CELL
 CC GROWTH (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 CC -1- INDUCTION: BY AUXIN.
 CC -1- SIMILARITY: BELONGS TO THE AUX/IAA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; U18408; AAC49047.1; -
 DR Mendel; 7122; Arath;1524;7122.
 DR InterPro: IPR003311; AUX_IAA.
 DR Pfam; PF02309; AUX_IAA; 1.
 KW Multigene family; Nuclear protein; Translation regulation.
 SQ SEQUENCE 189 AA; 21017 MW; 5E9C1477A3877430 CRC64;

Query Match 50.0%; Score 40; DB 1; Length 189;
 RT Best Local Similarity 58.3%; Pred. No. 4.2;
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 2 DQDQVVVQVPM 13
 ID 152 DRDMALVGVPM 163

RESULT 5
 ID ACSA_SYNY3 STANDARD; PRT; 653 AA.
 AC Q55404;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ACETYL-COENZYME A SYNTHETASE (EC 6.2.1.1) (ACETATE--COA LIGASE) (ACYL-
 DE ACTIVATING ENZYME).
 GN ACS OR SL0542.
 OS Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 OX NCBI_TaxID=1148;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96127529; PubMed=8590279;
 RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
 RA Sugita M., Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
 RT region from map positions 648 to 928 of the genome.";
 RL DNA Res. 2:153-166(1995).
 CC -1- CATALYTIC ACTIVITY: ATP + ACETATE + COA = AMP + PYROPHOSPHATE +
 CC ACETYL-COA.
 CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; D64003; BAA10498.1; -
 DR InterPro: IPR000873; AMP-bind.
 DR Pfam; PF00501; AMP-binding; 1.

DR PRINTS: PR00154; AMPBINDING.
DR PROSITE: PS00455; AMP_BINDING; 1.
KW Ligase; Complete proteome.
SQ SEQUENCE 653 AA; 73052 MW; 5EBCAD430EEFFA03 CRC64;

Query Match 50.0%; Score 40; DB 1; Length 653;
Best Local Similarity 46.2%; Pred. No. 16;
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 DDDQMVVQVMP 14
11 :|::|111
DB 464 DGGFTVIKQPM 476

RESULT 6
PHNJ_ECOLI STANDARD; PRT; 281 AA.
ID PHNJ_ECOLI
AC P16686;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PHNJ PROTEIN.
GN PHNJ OR B4098.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_Taxid=562;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=91193228; PubMed=1840580;
RA Makino K., Kim S.K., Shinagawa H., Amenura M., Nakata A.;
RT "Molecular analysis of the cryptic and functional phn operons for
phosphonate use in Escherichia coli K-12.";
RL J. Bacteriol. 173:2665-2672(1991).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN-K12 / MG1655;
RX MEDLINE=9534362; PubMed=7610040;
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
Blattner F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
region from 92.8 through 100 minutes.";
RL Nucleic Acids Res. 23:2105-2119(1995).
[3]
RN SEQUENCE FROM N.A.
RP STRAIN-B;
RX MEDLINE=90170953; PubMed=2155230;
RA Chen C.-M., Ye Q.-Z., Zhu Z., Wanner B.L., Walsh C.T.;
RT "Molecular biology of carbon-phosphorus bond cleavage. Cloning and
sequencing of the phn (psid) genes involved in alkylphosphonate
uptake and C-P lyase activity in Escherichia coli B.";
RL J. Biol. Chem. 265:4461-4471(1990).
-1- FUNCTION: BELONGS TO AN OPERON INVOLVED IN ALKYLPHOSPHONATE
UPTAKE AND C-P LYASE. EXACT FUNCTION NOT KNOWN.
-1- MISCELLANEOUS: THE SEQUENCE SHOWN IS THAT OF STRAIN K12.
-1- SIMILARITY: BELONGS TO THE PHNJ FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC
CC EMBL; D90227; BA14270.1; -
CC EMBL; U14003; AAA6997.1; -
CC EMBL; AE000482; AAC77059.1; -
CC EMBL; J05260; AAA24348.1; -
CC PIR; B35719; B35719.
CC Ecogene; EG10719; phnj.

KW Alkylphosphonate uptake; Complete proteome.
FT VARIANT 103 103 V -> L (IN STRAIN B).
SQ SEQUENCE 281 AA; 31845 MW; 241F6AFL40995468 CRC64;

Query Match 48.8%; Score 39; DB 1; Length 281;
Best Local Similarity 50.0%; Pred. No. 9.8;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 3 QDDQMVVQVMP 14
11 :|::|111
DB 117 EDQITIFQVPI 128

RESULT 7
YECC_ECOLI STANDARD; PRT; 188 AA.
ID YECC_ECOLI
AC P52007; P52008; P76295;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROTEIN YECC.
GN YECC OR B1875.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_Taxid=562;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
Maub B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN-K12;
RX MEDLINE=97251358; PubMed=9097040;
RA Itoh T., Alba H., Baba T., Fujita K., Hayashi K., Inada T.,
Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,
RA Saito N., Sempel G., Seki Y., Sivasubramanian S., Tagami H.,
RA Takeda J., Takemoto K., Wada C., Yamamoto Y., Horiuchi T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 40.1-50.0 min region on the linkage map.";
RL DNA Res. 3:379-392(1996).
[3]
RN SEQUENCE OF 102-188 FROM N.A.
RP STRAIN-K12 / MC4100;
RA Gupta S.D., Wu H.C.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
[4]
RN IDENTIFICATION BY MASS SPECTROMETRY.
RX MEDLINE=99420866; PubMed=10493123;
RA Fountoulakis M., Takacs M.-F., Berndt P., Langan H., Takacs B.;
RT "Enrichment of low abundance proteins of Escherichia coli by
hydroxyapatite chromatography.";
RL Electrophoresis 20:2181-2195(1999).
-1- SIMILARITY: STRONG, TO H-INFLUENZAE H11582/H11581.
-1- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
FRAMESHIFT IN POSITION 128 THAT PRODUCES TWO SEPARATE ORFS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC


```

CC -----
DR EMBL: AE000281; AAC74945.1; ALT_INIT.
DR EMBL: D900829; BAA15685.1; -.
DR EMBL: D900830; BAA15692.1; -.
DR EMBL: L38618; AAA89200.1; ALT_FRAME.
DR EMBL: L38618; AAA89201.1; ALT_FRAME.
DR Ecocore; E013405; yecm.
KM Complete proteome.
FT CONFLICT 102 G -> R (IN REF. 3).
FT CONFLICT 120 G -> V (IN REF. 3).
FT CONFLICT 167 G -> A (IN REF. 3).
SQ SEQUENCE 188 AA; 21205 MW; 700B8E042FMAA8A5 CRC64;

Query Match
Best Local Similarity 47.5%; Score 38; DB 1; Length 188;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 5 OMVWVQVMP 14
DB 92 QFISVLPMP 101

RESULT 8
ID DYR2_HUMAN STANDARD; PRT; 528 AA.
AC 092630;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DUAL-SPECIFICITY TYROSINE-PHOSPHORYLATION REGULATED KINASE 2
DE (EC 2.7.1.1.-).
GN DYRK2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98421512; PubMed=9748265;
RA Becker W., Weber Y., Wetzel K., Elmbjör K., Tejedor F.J.,
RA Joost H.-G.;
RT "Sequence characteristics, subcellular localization, and substrate
RT specificity of DYRK-related kinases, a novel family of dual
RT specificity protein kinases."
RT J. Biol. Chem. 273:25893-25902(1998).
RN [2]
RP SEQUENCE OF 320-528 FROM N.A.
RC TISSUE=Placenta;
RA Becker W., Joost H.-G.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: IN VITRO; CAN PHOSPHORYLATE HISTONES H3 AND H2B ON SER
CC AND THR RESIDUES. MAY BE INVOLVED IN THE REGULATION OF CELLULAR
CC GROWTH AND/OR DEVELOPMENT.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- PTM: AUTOPHOSPHORYLATED ON TYR RESIDUES.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC MNB/DYRK SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: Y13493; CAA13885.1; -.
DR EMBL: Y09216; CAA70418.1; -.
DR HSSP: Q16539; 1WFC.
DR MIM: 603496; -.
DR InterPro: IPR000719; Euk_pkinase.

```

```

DR InterPro: IPR002290; Ser_thr_kin_actsite.
DR Pfam: PF00069; pkinase; 2.
DR SMART: SMO0220; S_TKc; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
KM Transferrase: Serine/threonine-protein kinase; Tyrosine-protein kinase;
KM ATP-binding; phosphorylation.
FT DOMAIN 149 462
FT NP_BIND 155 163
FT BINDING 178 178
FT ACT_SITE 275 275
SQ SEQUENCE 528 AA; 59714 MW; AF2C6822ED9522D7 CRC64;

Query Match
Best Local Similarity 47.5%; Score 38; DB 1; Length 528;
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 YDODKMVVQVP 12
DB 130 YDDDGSTVQVP 141

RESULT 9
ID ERG1_HUMAN STANDARD; PRT; 574 AA.
AC Q14534; Q9UEK6;
DT 01-NOV-1997 (Rel. 35, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE SQUALENE MONOOXYGENASE (EC 1.14.99.7) (SQUALENE EPOXIDASE) (SE).
GN SQR OR ERG1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97432831; PubMed=9286711;
RA Nagai M., Sakakibara J., Makui K., Fukushima Y., Igarashi S.,
RA Tsuji S., Arakawa M., Ono T.;
RT "Localization of the squalene epoxidase gene (SQR) to human
RT chromosome region 8q24.1."
RT Genomics 44:141-143(1997).
RN [2]
RP SEQUENCE OF 187-535 FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=96215195; PubMed=8626488;
RA Nakamura Y., Sakakibara J., Izumi T., Shibata A., Ono T.;
RT "Transcriptional regulation of squalene epoxidase by sterols and
RT inhibitors in HeLa cells."
RL J. Biol. Chem. 271:8053-8056(1996).
CC -1- FUNCTION: CATALYZES THE FIRST OXYGENATION STEP IN STEROL
CC BIOSYNTHESIS AND IS SUGGESTED TO BE ONE OF THE RATE-LIMITING
CC ENZYMES IN THIS PATHWAY.
CC -1- CATALYTIC ACTIVITY: SQUALENE + AH(2) + O(2) - (S)-SQUALENE-2,3-
CC EPOXIDE + A + H(2)O.
CC -1- COFACTOR: FAD.
CC -1- SUBUNIT: MAY FORM A COMPLEX WITH SQUALENE SYNTHASE.
CC -1- SIMILARITY: BELONGS TO THE SQUALENE MONOOXYGENASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: D78130; BAA22372.1; -.
DR EMBL: D78129; BAA11209.1; -.

```

```

DR MIM; 602019; -
DR InterPro: IPR000205; NAD_binding.
DR InterPro: IPR003042; Rng_mnoxgase.
DR InterPro: IPR007733; flavo_mnoxgase.
DR Pfam: PF01360; Monooxygenase.1.
DR PRINTS: PR00420; RNMNOXGNASE.
KM Oxidoreductase; Flavoprotein; FAD; Transmembrane.
FT TRANSMEM 20 40
FT TRANSMEM 61 81
FT TRANSMEM 123 143
FT TRANSMEM 546 566
FT TRANSMEM 126 153
FT NP_BIND 247 247
FT CONFLICT 247 247
FT CONFLICT 336 336
FT CONFLICT 389 389
FT CONFLICT 451 451
FT CONFLICT 518 518
SQ SEQUENCE 574 AA; 63939 MW; DD4F95308864FD09 CRC64;

Query Match
Best Local Similarity 47.5%; Score 38; DB 1; Length 574;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 YDQDMVWVQVWP 14
:|:|:|:|:|:|
Db 198 HDQSKSEVQIYP 211

RESULT 10
AX11_ARATH STANDARD; PRT; 168 AA.
AC P49677; 023312;
DT 01-FEB-1996 (Rel. 33, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE AUXIN-RESPONSIVE PROTEIN IAA1 (INDOLEACETIC ACID-INDUCED PROTEIN 1).
OS IAA1 OR AT4G14560 OR DL3320W.
GN Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsis.
OC NCBI_Taxid=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA MEDLINE=20083488; PubMed=10617198;
RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
RA Harris B., Anstorge W., Brandt P., Grivell L.A., Rieger M.,
RA Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
RA Kreis M., Detseny M., Puigdomenech P., Watson M., Schmidheini T.,
RA Reichert B., Porteleille D., Perez-Alonso M., Bouty M., Bancroft I.,
RA Vos P., Hohelsel J., Zimmermann W., Wedler H., Ridley P.,
RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
RA Van der Schueren J., Gymonprez B., Chuang Y.-J., Vandenbussche F.,
RA Braeken M., Weljens I., Voet M., Bastlaens I., Aert R., Delcor E.,
RA Weitzenecker T., Bothe G., Rampsperger U., Hilbert H., Braun M.,
RA Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,
RA Moeljan P., Klein Lanhorst R., Rose M., Hauf J., Koetter P.,
RA Benelaiser S., Hempel S., Feldpausch M., Lambrecht S., Van den Daele H.,
RA De Keyser A., Buysmaert C., Gielen J., Villarroel R., De Clercq R.,
RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
RA Clark L., Doggett J., Hall S., Kay M., Leonard N., McIlroy K., Mayes R.,
RA Petrelet A., Rajendram M.-A., Lyne M., Benes V., Rechmann S.,
RA Borkova D., Blocker H., Scharfe M., Grimm M., Loehnert T.-H.,

```

```

RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fairman B., Granderath K., Danner D., Hetzl A.,
RA Neumann S., Argirion A., Vitale D., Liguori R., Piravandi E.,
RA Massenot O., Quigley F., Clabaud G., Muendlein A., Felber R.,
RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA Cherdorf F., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA Gibbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A.,
RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielke C.,
RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA Zaccaria P., Bayan M., Wilson R.K., de la Bastide M., Hebermann K.,
RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
RA Stoneking T., Kalkicki J., Graves T., Harton G., Edwards J.,
RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA Marx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
RA Antonolou B., Zidanic M., Strong C., Sun H., Lamar B., Jordan C.,
RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Tili S.,
RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
RA Chen E., Maira M., Martienssen R., McCombie W.R.;
RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
RT thaliana."
RL Nature 402:769-777(1999).
CC -1- FUNCTION: COULD ACT AS REGULATOR OF GENES RESPONSIBLE FOR
CC MEDIATING THE VARIOUS AUXIN-INDUCED EVENTS RESPONSIBLE FOR CELL
CC GROWTH.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- INDUCTION: BY AUXIN.
CC -1- SIMILARITY: BELONGS TO THE AUX/IAA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; L15448; AAA16569.1; -
CC DR EMBL; Z97336; CAB10235.1; -
CC DR EMBL; AL161539; CAB78498.1; -
CC DR Mendel; 25848; Arath:1524;25848.
CC InterPro: IPR003311; AUX_IAA.
CC Pfam; PF02309; AUX_IAA.1.
CC MultiGene Family; Nuclear protein; Translation regulation.
CC CONFLICT 52 52
CC FT A -> T (IN REF. 1).
CC SEQUENCE 168 AA; 19031 MW; 17CA9B3BB7962EE CRC64;

Query Match
Best Local Similarity 46.2%; Score 37; DB 1; Length 168;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 DQDMVWVQVWP 13
| | | | |
Db 135 DQDMVWVQVWP 146

RESULT 11
AX12_ARATH STANDARD; PRT; 174 AA.
AC P49678; 022596;
DT 01-FEB-1996 (Rel. 33, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE AUXIN-RESPONSIVE PROTEIN IAA2 (INDOLEACETIC ACID-INDUCED PROTEIN 2).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC eucots II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=94105161; PubMed=8278386;
 RA Abel S., Oeller P.W., Theologis A.;
 RT "Early auxin-induced genes encode short-lived nuclear proteins.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:326-330(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. MASSILEWSKIIA;
 RX "Molecular characterization of a genomic clone encoding the auxin-
 inducible IAA4 gene from Arabidopsis thaliana.";
 RL (in) Plant Gene Register PG97-178.
 CC -1- FUNCTION: COULD ACT AS REGULATOR OF GENES RESPONSIBLE FOR
 MEDIATING THE VARIOUS AUXIN-INDUCED EVENTS RESPONSIBLE FOR CELL
 GROWTH.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- INDUCTION: BY AUXIN.
 CC -1- SIMILARITY: BELONGS TO THE AUX/IAA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L15449; AAA16570.1; -
 DR EMBL: AF021157; AAB97164.1; -
 DR Mendel: 25258; Arabid.1524;25258.
 DR InterPro: IPR003311; AUX_IAA.
 DR Pfam: PF02309; AUX_IAA; 1.
 KW Multigene family; Nuclear protein; Translation regulation.
 FT CONFLICT 38 I -> S (YN REF. 1)
 FT SEQUENCE 174 AA: 19936 MW: D2E233MBF0107DB8 CRC64;
 SQ
 Query Match 46.2%; Score 37; DB 1; Length 174;
 Best Local Similarity 58.3%; Pred. No. 13;
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 QY 2 DDDQWVVOVPM 13
 DB 138 DGDWMLVGDVPM 149
 ID AX16_PEA STANDARD: PRT; 179 AA.
 AC P49680;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE AUXIN-INDUCED PROTEIN IAA6.
 GN IAA6.
 OS Pisum sativum (Garden pea).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC eucots II; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.
 OX NCBI_TaxID=3888;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. ALASKA;
 RX MEDLINE=94016594; PubMed=8411182;
 RA Oeller P.W., Keller J.A., Parks J.E., Silbert J.E., Theologis A.;
 RT "Structural characterization of the early indoleacetic acid-inducible
 genes, PS-IAA4/5 and PS-IAA6, of pea (Pisum sativum L.).";
 RL J. Mol. Biol. 233:789-798(1993).

CC -1- FUNCTION: COULD ACT AS REGULATOR OF GENES RESPONSIBLE FOR
 MEDIATING THE VARIOUS AUXIN-INDUCED EVENTS RESPONSIBLE FOR CELL
 GROWTH (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- INDUCTION: BY AUXIN.
 CC -1- SIMILARITY: BELONGS TO THE AUX/IAA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X68218; CAA48300.1; -
 DR EMBL: X68217; CAA48299.1; -
 DR InterPro: IPR003311; AUX_IAA.
 DR Pfam: PF02309; AUX_IAA; 1.
 KW Multigene family; Nuclear protein; Translation regulation.
 FT SEQUENCE 179 AA: 20330 MW: 0D212F384D100D7E CRC64;
 SQ
 Query Match 46.2%; Score 37; DB 1; Length 179;
 Best Local Similarity 58.3%; Pred. No. 14;
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 QY 2 DDDQWVVOVPM 13
 DB 135 DGDWMLVGDVPM 146
 ID AX14_ARATH STANDARD: PRT; 186 AA.
 AC P33077;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE AUXIN-RESPONSIVE PROTEIN IAA4 (INDOLEACETIC ACID-INDUCED PROTEIN 4)
 DE (AUXIN-INDUCED PROTEIN AUX2-11).
 GN IAA4 OR AUX2-11.
 OS Arabidopsis thaliana (Mouse-ear cress).
 CC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC eucots II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=9138700; PubMed=2102379;
 RA Conner T.W., Goekjian V.H., Lafayette P.R., Key J.L.;
 RT "Structure and expression of two auxin-inducible genes from
 Arabidopsis.";
 RL Plant Mol. Biol. 15:623-632(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=94105161; PubMed=8278386;
 RA Abel S., Oeller P.W., Theologis A.;
 RT "Early auxin-induced genes encode short-lived nuclear proteins.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:326-330(1994).
 CC -1- FUNCTION: COULD ACT AS REGULATOR OF GENES RESPONSIBLE FOR
 MEDIATING THE VARIOUS AUXIN-INDUCED EVENTS RESPONSIBLE FOR CELL
 GROWTH.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- INDUCTION: BY AUXIN.
 CC -1- SIMILARITY: BELONGS TO THE AUX/IAA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL; X53435; CAA37526.1; -.
DR EMBL; L15450; AAA16571.1; -.
DR PIR; S12243; S12243.
DR InterPro; IPR003311; AUX_1AA.
DR Pfam; PF02309; AUX_1AA; 1.
KM Multigene family; Nuclear protein; Translation regulation.
SQ SEQUENCE 186 AA; 20975 MW; DCF66CAC4D28C03B CRC64;

Query Match 46.2%; Score 37; DB 1; Length 186;
Best Local Similarity 58.3%; Pred. No. 14;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 DQDMVVVQVPM 13
Db 149 DGDMMVLGVDPV 160

RESULT 14
AX2C_PHAU STANDARD; PRT; 188 AA.
AC 024541;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE AUXIN-INDUCED PROTEIN 22C (INDOLE-3-ACETIC ACID INDUCED PROTEIN
DE ARG12).
GN AUX22C OR ARG12.
OS Phaseolus aureus (Mung bean) (Vigna radiata).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Vigna.
OX NCBI_TaxID=3916;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Hypocotyl;
RA Hashimoto H., Yamamoto K.T.;
RT "Three more members of the AUX/IAA gene family from mung bean (Vigna
RT radiata) hypocotyl."
RL (in) Plant Gene Register PGR97-137.
CC -1- FUNCTION: COULD ACT AS REGULATOR OF GENES RESPONSIBLE FOR
CC MEDIATING THE VARIOUS AUXIN-INDUCED EVENTS RESPONSIBLE FOR CELL
CC GROWTH (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- INDUCTION: BY AUXIN.
CC -1- SIMILARITY: BELONGS TO THE AUX/IAA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

Query Match 46.2%; Score 37; DB 1; Length 186;
Best Local Similarity 58.3%; Pred. No. 15;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 DQDMVVVQVPM 13
Db 139 DGDMMVLGVDPV 150

RESULT 15

AX14_PEA STANDARD; PRT; 189 AA.
AC P49679;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE AUXIN-INDUCED PROTEIN IAA4.
GN IAA4/5.
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.
OX NCBI_TaxID=3888;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, ALASKA;
RX MEDLINE=94016594; PubMed=8411182;
RA Oeller P.W., Keller J.A., Parks J.E., Silbert J.E., Theologis A.;
RT "Structural characterization of the early indoleacetic acid-inducible
RT genes, PS-IAA4/5 and PS-IAA6, of pea (Pisum sativum L.).";
RL J. Mol. Biol. 233:789-798(1993).
CC -1- FUNCTION: COULD ACT AS REGULATOR OF GENES RESPONSIBLE FOR CELL
CC GROWTH (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- INDUCTION: BY AUXIN.
CC -1- SIMILARITY: BELONGS TO THE AUX/IAA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

Query Match 46.2%; Score 37; DB 1; Length 189;
Best Local Similarity 58.3%; Pred. No. 15;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 DQDMVVVQVPM 13
Db 153 DGDMMVLGVDPV 164

Search completed: November 30, 2001, 09:40:56
Job time: 35 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OK protein - protein search, using sw model

Run on: November 30, 2001, 09:44:54 ; Search time 67.45 Seconds
(without alignments)
30.360 Million cell updates/sec

Title: US-08-846-234-3
Perfect score: 80
Sequence: 1 YPDQDMVVVQVWMP 14

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP invertebrate:*
6: SP mammal:*
7: SP mnc:*
8: SP organelle:*
9: SP phage:*
10: SP plant:*
11: SP rodent:*
12: SP virus:*
13: SP vertebrate:*
14: SP unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	80	100.0	784	10	Q92T62
2	49	61.2	783	10	Q9FND9
3	45	56.2	570	10	Q9LDE3
4	44	55.0	164	10	Q38832
5	44	55.0	228	10	Q9C5W8
6	44	55.0	229	10	Q49162
7	44	55.0	234	10	Q23311
8	44	55.0	619	10	Q9XID4
9	44	55.0	719	10	Q9LYM1
10	43	53.8	423	10	Q81350
11	42	52.5	500	10	Q9X178
12	42	52.5	529	12	Q39281
13	42	52.5	600	10	Q9C8N7
14	42	52.5	767	10	Q9LOE8
15	40	50.0	189	10	Q9C539
16	40	50.0	202	10	Q9XEX3
17	40	50.0	847	2	Q9ADT8
18	40	50.0	981	5	Q9NN63
19	40	50.0	1159	4	Q9UF39

20	39.5	49.4	345	7	Q19480	O19480 heterodontu
21	39	48.8	93	5	Q9U194	Q9U194 leishmania
22	39	48.8	206	5	Q00629	Q00629 manduca sex
23	39	48.8	288	5	Q9XZD6	Q9XZD6 plasmodium
24	39	48.8	305	10	Q9ZS18	Q9ZS18 arabidopsis
25	39	48.8	344	3	Q9C2J0	Q9C2J0 neurospora
26	39	48.8	384	2	Q9PD55	Q9PD55 xyella fas
27	39	48.8	565	12	Q9PY76	Q9PY76 human calic
28	39	48.8	848	13	Q9XGV9	Q9XGV9 anguilla ja
29	39	48.8	2435	5	Q9W433	Q9W433 drosophila
30	38	47.5	73	4	Q9HC65	Q9HC65 homo sapien
31	38	47.5	138	2	Q9FCE9	Q9FCE9 streptomyce
32	38	47.5	162	2	Q9PFC0	Q9PFC0 xyella fas
33	38	47.5	175	10	Q9XE14	Q9XE14 oryza sativ
34	38	47.5	390	2	P96217	P96217 mycobacteri
35	38	47.5	390	12	Q9WRN9	Q9WRN9 macaca mula
36	38	47.5	402	5	Q9V6P0	Q9V6P0 drosophila
37	38	47.5	404	11	Q54929	Q54929 mus musculu
38	38	47.5	404	11	Q9RLM8	Q9RLM8 mus musculu
39	38	47.5	473	10	Q9SHZ2	Q9SHZ2 arabidopsis
40	38	47.5	486	2	Q30731	Q30731 rhodobacter
41	38	47.5	522	12	Q9WHJ1	Q9WHJ1 walleye epi
42	38	47.5	524	2	Q9KSI8	Q9KSI8 vibrio chol
43	38	47.5	569	2	Q69545	Q69545 mycobacteri
44	38	47.5	574	4	Q9UNR6	Q9UNR6 homo sapien
45	38	47.5	589	10	Q9M310	Q9M310 elicer arlet

ALIGNMENTS

RESULT	ID	Query Match	Score	DB ID	Length	Best Local Similarity	Pred. No.	Mismatches	Indels	Gaps
1	Q92T62	100.0%	80	10	784	100.0%	9.4e-06	0	0	0
2	Q9FND9	100.0%	80	10	783	100.0%	9.4e-06	0	0	0
3	Q9LDE3	100.0%	80	10	570	100.0%	9.4e-06	0	0	0
4	Q38832	100.0%	80	10	164	100.0%	9.4e-06	0	0	0
5	Q9C5W8	100.0%	80	10	228	100.0%	9.4e-06	0	0	0
6	Q49162	100.0%	80	10	229	100.0%	9.4e-06	0	0	0
7	Q23311	100.0%	80	10	234	100.0%	9.4e-06	0	0	0
8	Q9XID4	100.0%	80	10	619	100.0%	9.4e-06	0	0	0
9	Q9LYM1	100.0%	80	10	719	100.0%	9.4e-06	0	0	0
10	Q81350	100.0%	80	10	423	100.0%	9.4e-06	0	0	0
11	Q9X178	100.0%	80	10	500	100.0%	9.4e-06	0	0	0
12	Q39281	100.0%	80	12	529	100.0%	9.4e-06	0	0	0
13	Q9C8N7	100.0%	80	10	600	100.0%	9.4e-06	0	0	0
14	Q9LOE8	100.0%	80	10	767	100.0%	9.4e-06	0	0	0
15	Q9C539	100.0%	80	10	189	100.0%	9.4e-06	0	0	0
16	Q9XEX3	100.0%	80	10	202	100.0%	9.4e-06	0	0	0
17	Q9ADT8	100.0%	80	2	847	100.0%	9.4e-06	0	0	0
18	Q9NN63	100.0%	80	5	981	100.0%	9.4e-06	0	0	0
19	Q9UF39	100.0%	80	4	1159	100.0%	9.4e-06	0	0	0

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=COLUMBIA;
RC MEDLINE=98069011; PubMed=9405937;
RA Kotani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. II.
RT Sequence features of the regions of 1,044,062 bp covered by thirteen
RT physically assigned P1 clones."
RL DNA Res. 4:291-300(1997).
DR EMBL: AB006702; BAB11595.1;
SQ SEQUENCE 783 AA; 86237 MW; 3C37D1D7871888AC CRC64;

Query Match 61.2%; Score 49; DB 10; Length 783;
Best Local Similarity 72.7%; Pred. No. 2.8;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 DODQMVVQVWP 13
DB 760 EDSMVVQVWP 770

RESULT 3
ID 09LOE3 PRELIMINARY; PRT; 570 AA.
AC 09LOE3;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE F1504.42.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
[1]
RN SEQUENCE FROM N.A.
RP Ecker J.R.;
RA Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RP Ecker J.R.;
RA Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A.
RP Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
RA Khan S., Kim C., Altati H., Bel B., Chin C., Chlou J., Choi E.,
RA Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,
RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharbeky N.,
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
RA Thavari A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
RA Theologis A., Ecker J.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC007887; AAF79360.1;
DR InterPro: IPR003340; B3.
DR Pfam: PF02362; B3; 1.
SQ SEQUENCE 570 AA; 65041 MW; E64A987BA713DF7E CRC64;

Query Match 56.2%; Score 45; DB 10; Length 570;
Best Local Similarity 53.8%; Pred. No. 10;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 DODQMVVQVWP 14
DB 558 DEDMLVGDVPWP 570

RESULT 4
ID 038832 PRELIMINARY; PRT; 164 AA.
AC 038832;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE AUXIN-INDUCIBLE IAA14 (FRAGMENT).
GN IAA14.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=95387393; PubMed=7658471;
RA Abel S., Nguyen M.D., Theologis A.;
RT "The PS-IAA4/5-like family of early auxin-inducible mRNAs in
RT Arabidopsis thaliana".
RL J. Mol. Biol. 251:533-549(1995).
DR EMBL: U18416; AAC49055.1;
DR Mendel: 7130; Arch:1524;7130.
DR InterPro: IPR003311; AUX_1AA.
DR Pfam: PF02309; AUX_1AA; 1.
FT NON_TER
SQ SEQUENCE 164 AA; 18232 MW; D7BB6B71F6DD3574 CRC64;

Query Match 55.0%; Score 44; DB 10; Length 164;
Best Local Similarity 61.5%; Pred. No. 3.9;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 DODQMVVQVWP 14
DB 120 DGDWMLVGDVPWP 132

RESULT 5
ID 09C5W8 PRELIMINARY; PRT; 228 AA.
AC 09C5W8;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE IAA14.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
[1]
RN SEQUENCE FROM N.A.
RP Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Goldsmith A.D., Lee J.M., Toriumi M., Yu G., Brooks S., Chao Q.,
RA Chen H., Karlin-Neumann G., Kim C., Lam B., Miranda M.,
RA Palm C.J., Shinn P., Southwick A., Davis R.W., Ecker J.R.,
RA Theologis A.;
RL "Full length cDNA of gene d13315c/AT4g14550 (GI:7268161)."
RT Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF34718; AAG50096.1;
DR InterPro: IPR003311; AUX_1AA.
DR Pfam: PF02309; AUX_1AA; 1.
SQ SEQUENCE 228 AA; 25044 MW; A76A9F733538E372 CRC64;

Query Match 55.0%; Score 44; DB 10; Length 228;
Best Local Similarity 61.5%; Pred. No. 5.6;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 DODQMVVQVWP 14
DB 184 DGDWMLVGDVPWP 196


```

RESULT 6
ID 049162 PRELIMINARY; PRT; 229 AA.
AC 049162;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE IAA17/AXR3-1 PROTEIN.
GN AXR3.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=98146427; PubMed=9478901;
RA Rouse D., Mackay P., Stirnberg P., Estelle M., Leyser O.;
RT "Changes in auxin response from mutations in an AUX/IAA gene.";
RL Science 279:1371-1373(1998).
DR EMBL; AF040632; AAC39440.1; -.
DR Mendel; 24578; Arabid.1524;24578.
DR InterPro; IPR003311; AUX_IAA.
DR Pfam; PF02309; AUX_IAA; 1.
SQ SEQUENCE 229 AA; 25304 MW; BDAA393475393545 CRC64;

Query Match 55.0%; Score 44; DB 10; Length 229;
Best Local Similarity 61.5%; Pred. No. 5.6;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 2 DDDQWVGVVWVMP 14
DB 185 DGDWMLVGDVMP 197

RESULT 7
ID 023311 PRELIMINARY; PRT; 234 AA.
AC 023311;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE IAA7-LIKE PROTEIN (IAA7 LIKE PROTEIN).
GN DL315C OR ATG14550.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Stiekema W., Murphy G., Wambutt R., Pohl T., Terry N.,
RA Kreis M., Kavanagh T., Entian K.D., Kleger W., James R.,
RA Pulgadenhoch P., Hatzipoulos P., Obermaier B., Duesterhoft A.,
RA Jones J., Palme K., Ansoerge W., Delseny M., Bancroft I., Mewes H.W.,
RA Schueller C., Chaitatzis N.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z97336; CAB46059.1; -.
DR EMBL; AL161539; CAB78497.1; -.
DR Mendel; 25847; Arabid.1524;25847.
DR InterPro; IPR003311; AUX_IAA.
DR Pfam; PF02309; AUX_IAA; 1.
SQ SEQUENCE 234 AA; 25667 MW; 340A94E95B23FB60 CRC64;

```

```

Query Match 55.0%; Score 44; DB 10; Length 234;
Best Local Similarity 61.5%; Pred. No. 5.7;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 2 DDDQWVGVVWVMP 14
DB 184 DGDWMLVGDVMP 196

RESULT 8
ID 09XID4 PRELIMINARY; PRT; 619 AA.
AC 09XID4;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 70.0 KDA PROTEIN F23M19.4.
GN F23M19.4.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,
RA Lee J., Liu A., Li J., Kremenetskaia I., Lutos J., Gonzalez A.,
RA Altafi H., Araujo R., Chao Q., Conn L., Conway A.B., Dunn P.,
RA Hansen N., Huizar L., Kim C., Palm C., Rowley D., Shinn P., Walker M.,
RA Davis R.W., Ecker J.R., Federapfel N.A., Theologis A.;
RT "Arabidopsis thaliana chromosome 1 BAC F23M19 sequence.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Theologis A.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC007434; AAD39615.1; -.
DR InterPro; IPR003311; AUX_IAA.
DR InterPro; IPR003340; B3.
DR Pfam; PF02309; AUX_IAA; 1.
DR Pfam; PF02362; B3; 1.
KW Hypothetical protein.
SQ SEQUENCE 619 AA; 69994 MW; 1DA4DD7BC43927E7 CRC64;

Query Match 55.0%; Score 44; DB 10; Length 619;
Best Local Similarity 53.8%; Pred. No. 17;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 2 DDDQWVGVVWVMP 14
DB 592 DEDKMLVGDVMP 604

RESULT 9
ID 09LYN1 PRELIMINARY; PRT; 719 AA.
AC 09LYN1;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 77.3 KDA PROTEIN.
GN F18021_60.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Benes V., Wurnbach E., Drzonek H., Ansoerge W., Mewes H.W., Rudd S.,

```

RA Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL163763; CAB87409.1; -
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR003592; LRR_out.
DR Pfam; PR000560; LRR; 5.
DR Pfam; PR00069; pkinase; 2.
DR PRINTS; PR00019; LEURICHRPT.
DR SMART; SM00370; LRR; 6.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR ATP-binding; Hypothetical protein; Transferase.
KW SEQUENCE 719 AA; 77329 MW; 3B91891A64E0D9D3 CRC64;
SQ

Query Match 55.0%; Score 44; DB 10; Length 719;
Best Local Similarity 54.5%; Pred. No. 20;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 3 DODMNVVQVPM 13
DB 100 QGOVIVIQLPW 110

RESULT 10
O81350 PRELIMINARY; PRT; 423 AA.
AC O81350;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE 5'-ADENYLYLSULFATE REDUCTASE (EC 1.8.99.).
OS Enteromorpha intestinalis (Hollow green seaweed).
OC Eukaryota; Viridiplantae; Chlorophyta; Ulvophyceae; Ulvales; Ulvaceae;
OC Enteromorpha.
OX NCBI_TaxID=3116;
RN [1]
RP SEQUENCE FROM N.A.
RA Geo Y., Lestek T.;
RT "Cloning of 5'-adenylylsulfate (APS) reductase from the marine
RT macroalgae Enteromorpha intestinalis.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF069951; AAC26855.1; -
DR Mendel; 31988; Entin;1035;31988.
DR InterPro; IPR000063; ThioRed.
DR InterPro; IPR002500; PAPS_reduct.
DR Pfam; PF00085; thioRed; 1.
DR Pfam; PF01507; PAPS_reduct; 1.
DR PRINTS; PR00421; THIOREDOXIN.
DR Oxidoreductase.
KW SEQUENCE 423 AA; 46808 MW; 808BE4F3BD56F1F2 CRC64;
SQ

Query Match 53.8%; Score 43; DB 10; Length 423;
Best Local Similarity 58.3%; Pred. No. 17;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 2 DODMNVVQVPM 13
DB 330 DKDMNVVYAPW 341

RESULT 11
O9X178 PRELIMINARY; PRT; 500 AA.
AC O9X178;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)

DE F7A19.18 PROTEIN.
GN F7A19.18.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
RA Altafi H., Araujo R., Huitzer L., Rowley D., Buehler E., Dunn P.,
RA Gonzalez A., Kremenekskaja I., Kim C., Lenz C., Li J., Liu S.,
RA Lueros S., Schwartz J., Shinn P., Toriumi M., Vysotskaia V.S.,
RA Walker M., Yu G., Ecker J., Theologis A., Davis R.W.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC007576; AAD39294.1; -
KW SEQUENCE 500 AA; 57538 MW; 1C5BB6152BFAE690 CRC64;
SQ

Query Match 52.5%; Score 42; DB 10; Length 500;
Best Local Similarity 50.0%; Pred. No. 30;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 2 DODMNVVQVPM 13
DB 225 ESDQILIRQVPM 236

RESULT 12
O39281 PRELIMINARY; PRT; 529 AA.
AC O39281;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE COUNTERPART OF HSV-1 GENE UL21 AND VZV GENE 38.
GN 40.
OS Equine herpesvirus 4.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicelloviruses.
OX NCBI_TaxID=10331;
RN [1]
RP SEQUENCE OF 489-529 FROM N.A.
RC STRAIN=NS80567;
RC MEDLINE=90362066; PubMed=2167933;
RA Nicolson L., Cullinane A.A., Onions D.E.;
RT "The nucleotide sequence of an equine herpesvirus 4 gene homologue of
RT the herpes simplex virus 1 glycoprotein H gene.";
RL J. Gen. Virol. 71:0-0(0).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NS80567;
RC MEDLINE=98264497; PubMed=9603335;
RA Telford E.A., Watson M.S., Perry J., Cullinane A.A., Davison A.J.;
RT "The DNA sequence of equine herpesvirus-4.";
RL J. Gen. Virol. 79:1197-1203(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=NS80567;
RA Telford E.A., Watson M.S., Perry J., Cullinane A.A., Davison A.J.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF030027; AAC99558.1; -
KW SEQUENCE 529 AA; 58150 MW; E2B737B89D489413 CRC64;
SQ

Query Match 52.5%; Score 42; DB 12; Length 529;
Best Local Similarity 45.5%; Pred. No. 32;
Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

OY 4 DODMNVVQVPM 14
DB 283 DEILVEPWP 293

RESULT 13
 OC8N7 PRELIMINARY: PRT: 600 AA.
 ID 09C8N7
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE AUXIN RESPONSE FACTOR, PUTATIVE.
 GN F7P12.6.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eucots II; Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_TaxID=3702;
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=21016719; PubMed=11130712;
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
 White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
 Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
 Chung M.K., Conn L., Conway A.B., Creasy T.H., Dewar K.,
 Dunn P., Egu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
 Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hultzer L.,
 Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
 Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
 Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 Lin X., Liu S.X., Liu Z.A., Luros J.S., Malti R., Marzilli A.,
 Miltischer J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
 Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
 Uterbeck T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
 Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.,
 *Sequence and analysis of chromosome 1 of the plant Arabidopsis
 thaliana.*
 RT Nature 408:816-820(2000).
 RL EMBL: AC023913; AAG51894.1; -
 DR InterPro: IPR003311; AUX_1AA.
 DR InterPro: IPR003340; B3.
 DR Pfam: PF02309; AUX_1AA; 1.
 DR Pfam: PF02362; B3; 1.
 SO SEQUENCE 600 AA; 67933 MW; C28A2B4E79073FFF CRC64;

Query Match 52.5%; Score 42; DB 10; Length 600;
 Best Local Similarity 53.8%; Pred. No. 36;
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 DDDQWVGVQVMP 14
 DB 566 DDDKMLVGDMP 578

RESULT 14
 Q9LQ8 PRELIMINARY: PRT: 767 AA.
 ID 09LQ8
 AC 09LQ8;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE F1504.37.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eucots II; Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_TaxID=3702;
 RN
 RP SEQUENCE FROM N.A.
 RA Ecker J.R.;
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.

RN [2]
 RP SEQUENCE FROM N.A.
 RA Ecker J.R.;
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
 Khan S., Kim C., Altafi H., Bel B., Chin C., Chou J., Choi E.,
 Conn L., Conway A., Gonzalez A., Hansen N., Howling B., Koo T., Lam B.,
 Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharsky N.,
 Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
 Theaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
 RA Theologis A., Ecker J.;
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: TO GTP-BINDING ELONGATION FACTOR FAMILY.
 DR EMBL: AC007887; AAF79371.1; -
 DR InterPro: IPR003340; B3.
 DR InterPro: IPR00795; GTP_EFTU.
 DR Pfam: PF02362; B3; 1.
 DR Pfam: PF00009; GTP_EFTU; 1.
 KW GTP-binding; Protein biosynthesis
 SO SEQUENCE 767 AA; 85951 MW; FE29067F07D2E9F7 CRC64;

Query Match 52.5%; Score 42; DB 10; Length 767;
 Best Local Similarity 46.2%; Pred. No. 48;
 Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 DDDQWVGVQVMP 14
 DB 621 EDDKMLVGDMP 633

RESULT 15
 Q9C539 PRELIMINARY: PRT: 189 AA.
 ID 09C539
 AC 09C539;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE IAA6 (PUTATIVE IAA6 PROTEIN).
 GN F14G24.10.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eucots II; Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_TaxID=3702;
 RN
 RP SEQUENCE FROM N.A.
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
 Goldsmith A.D., Lee J.M., Toriumi M., Yu G., Brooks S., Chao Q.,
 Chen H., Karlin-Neumann G., Kim C., Lam B., Miranda M., Nguyen M.,
 Palm C.J., Shinn P., Southwick A., Davis R.W., Ecker J.R.,
 RA Theologis A.;
 RT *Full length cDNA of gene F14G24.10 (GI:6862944).*
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=21016719; PubMed=11130712;
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
 White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
 Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
 Chung M.K., Conn L., Conway A.B., Creasy T.H., Dewar K.,
 Dunn P., Egu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
 Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hultzer L.,
 Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
 Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
 Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 Lin X., Liu S.X., Liu Z.A., Luros J.S., Malti R., Marzilli A.,
 Miltischer J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
 Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,

RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
 RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
 RT *Sequence and analysis of chromosome 1 of the plant Arabidopsis
 RT thaliana.*;
 RT Nature 408:816-820(2000).
 DR EMBL; AF336915; AAG53996.1; -;
 DR EMBL; AC019018; AAG52268.1; -;
 DR InterPro; IPR003311; AUX_1AA.
 DR Pfam: PR02309; AUX_1AA; 1.
 SQ SEQUENCE 189 AA; 21031 MW; 5E9B130584A75465 CRC64;

Query Match 50.0%; Score 40; DB 10; Length 189;
 Best local Similarity 58.3%; Pred. No. 23;
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 DQDQWVYVQVPM 13
 I:I I:I I:I
 DB 152 DRDWMVLGQVPM 163

Search completed: November 30, 2001, 09:44:56
 Job time: 275 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 30, 2001, 09:43:38 ; Search time 34.66 Seconds
(without alignments)
9.090 Million cell updates/sec

Title: US-08-846-234-3

Perfect score: 80

Sequence: 1 YDQDQWVWVQVMP 14

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCVUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfiles1.pep:*

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	80	100.0	14	US-08-846-234-3	Sequence 3, Appl1
2	80	100.0	784	US-08-846-234-5	Sequence 5, Appl1
3	38	47.5	528	US-08-802-466-2	Sequence 2, Appl1
4	38	47.5	528	US-09-350-484-2	Sequence 2, Appl1
5	38	47.5	1829	US-09-157-420-1	Sequence 1, Appl1
6	37	46.2	156	US-09-064-922-2	Sequence 2, Appl1
7	37	46.2	179	US-09-064-922-5	Sequence 5, Appl1
8	37	46.2	399	US-09-491-362-7	Sequence 7, Appl1
9	37	46.2	452	US-08-764-870-16	Sequence 16, Appl1
10	37	46.2	452	US-08-980-115-16	Sequence 16, Appl1
11	37	46.2	475	US-09-449-362-2	Sequence 2, Appl1
12	37	46.2	477	US-09-449-335-6	Sequence 6, Appl1
13	37	46.2	477	US-09-449-335-6	Sequence 6, Appl1
14	37	46.2	918	US-09-041-886-11	Sequence 11, Appl1
15	36	45.0	400	US-09-184-001-4	Sequence 4, Appl1
16	36	45.0	404	US-09-184-001-2	Sequence 2, Appl1
17	36	45.0	3170	US-09-036-987A-4	Sequence 4, Appl1
18	36	45.0	3170	US-09-370-700-4	Sequence 4, Appl1
19	35	43.8	18	US-08-484-192-85	Sequence 85, Appl1
20	35	43.8	473	US-08-597-236-13	Sequence 13, Appl1
21	35	43.8	473	US-08-746-682A-13	Sequence 13, Appl1
22	35	43.8	589	US-07-668-648-6	Sequence 6, Appl1
23	35	43.8	589	US-08-429-998-6	Sequence 6, Appl1
24	35	43.8	589	US-08-431-333-6	Sequence 6, Appl1
25	35	43.8	589	PCT-US91-02321-6	Sequence 6, Appl1
26	35	43.8	1385	US-07-876-280-2	Sequence 2, Appl1
27	35	43.8	1385	US-07-675-772-2	Sequence 2, Appl1

28	35	43.8	1385	1	US-08-063-170-2	Sequence 2, Appl1
29	35	43.8	1385	1	US-08-158-232-2	Sequence 2, Appl1
30	35	43.8	1385	1	US-08-304-626-2	Sequence 2, Appl1
31	35	43.8	1385	1	US-08-316-301A-2	Sequence 2, Appl1
32	35	43.8	1385	2	US-08-611-928-2	Sequence 2, Appl1
33	35	43.8	1385	2	US-09-173-891-2	Sequence 2, Appl1
34	35	43.8	1385	4	US-09-076-137-2	Sequence 2, Appl1
35	35	43.8	1385	5	PCT-US92-03624-2	Sequence 2, Appl1
36	35	43.8	1385	6	5281530-1	Patent No. 5281530
37	35	43.8	1385	6	5426049-1	Patent No. 5426049
38	35	43.8	1642	1	US-08-447-411-45	Sequence 45, Appl1
39	35	43.8	1642	2	US-08-662-227-2	Sequence 2, Appl1
40	35	43.8	1642	4	US-09-017-947-2	Sequence 2, Appl1
41	35	43.8	1648	2	US-08-662-227-35	Sequence 35, Appl1
42	35	43.8	1648	4	US-09-017-947-35	Sequence 35, Appl1
43	35	43.8	2476	2	US-08-276-967-2	Sequence 2, Appl1
44	34	42.5	43	4	US-09-230-637-61	Sequence 61, Appl1
45	34	42.5	288	1	US-08-396-650-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1
US-08-846-234-3
; Sequence 3, Application US/08846234
; Patent No. 6166292
; GENERAL INFORMATION:
; APPLICANT: OSUMI Chleko
; APPLICANT: NOZAKI Jinshi
; APPLICANT: KIDA Takao
; TITLE OF INVENTION: RAFFINOSE SYNTHASE GENE, METHOD FOR
; TITLE OF INVENTION: PRODUCING RAFFINOSE, AND TRANSGENIC PLANT
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT, P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,234
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)-413-3000
; TELEFAX: (703)-413-2220
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: Internal
; US-08-846-234-3

Query Match 100.0%, Score 80; DB 4; Length 14;
Best Local Similarity 100.0%, Pred. No. 4.4e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YDQDQWVWVQVMP 14
DB 1 YDQDQWVWVQVMP 14

```
RESULT 2
US-08-846-234-5
; Sequence 5, Application US/08846234
; Patent No. 6166292
; GENERAL INFORMATION:
; APPLICANT: OSUMI Chieko
; APPLICANT: KIDA Takao
; TITLE OF INVENTION: RAFFINOSE SYNTHASE GENE, METHOD FOR
; TITLE OF INVENTION: PRODUCING RAFFINOSE, AND TRANSGENIC PLANT
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,234
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)-413-3000
; TELEFAX: (703)-413-2220
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 784 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-846-234-5

Query Match 100.0%; Score 80; DB 4; Length 784;
Best local Similarity 100.0%; Pred. No. 4.9e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YDDQDMVVOVWP 14
Db 756 YDDQDMVVOVWP 769

RESULT 3
US-08-802-466-2
; Sequence 2, Application US/08802466
; Patent No. 5972606
; GENERAL INFORMATION:
; APPLICANT: Creasy, et al.
; TITLE OF INVENTION: HUMAN PROTEIN KINASE HOACF72
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/802,466
```

```
; FILING DATE: 19 February 1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: William T. Han
; REGISTRATION NUMBER: 34,344
; REFERENCE/DOCKET NUMBER: GH50002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5219
; TELEFAX: 610-270-4026
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 528 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-802-466-2

Query Match 47.5%; Score 38; DB 2; Length 528;
Best local Similarity 66.7%; Pred. No. 64;
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YDDQDMVVOVP 12
Db 130 YDDQDMVVOVP 141

RESULT 4
US-09-350-484-2
; Sequence 2, Application US/09350484
; Patent No. 6159716
; GENERAL INFORMATION:
; APPLICANT: Creasy, et al.
; TITLE OF INVENTION: HUMAN PROTEIN KINASE HOACF72
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/350,484
; FILING DATE: 09-Jul-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/802,466
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: William T. Han
; REGISTRATION NUMBER: 34,344
; REFERENCE/DOCKET NUMBER: GH50002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5219
; TELEFAX: 610-270-4026
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 528 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
```

MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2
US-09-350-484-2

Query Match 47.5%; Score 38; DB 4; Length 528;
Best Local Similarity 66.7%; Pred. No. 64;
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YDODMNVVQVP 12
DB 130 YDDDGSGYVQVP 141

RESULT 5
US-09-157-420-1
Sequence 1, Application US/09157420

PATENT NO. 6180760
GENERAL INFORMATION:
APPLICANT: TAKAI, Yoshiaki
APPLICANT: NAKANISHI, Hiroyuki
APPLICANT: MANDAI, Kenji
APPLICANT: WADA, Manabu
APPLICANT: OBAISHI, Hiroshi
TITLE OF INVENTION: ACTIN FILAMENT-BINDING PROTEIN "L-AFADIN"
FILE REFERENCE: 98-1042*/LC(WMC)/653
CURRENT APPLICATION NUMBER: US/09157,420
CURRENT FILING DATE: 1998-09-21
NUMBER OF SEQ ID NOS: 1
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 1829
TYPE: PRT
ORGANISM: rat
US-09-157-420-1

Query Match 47.5%; Score 38; DB 4; Length 1829;
Best Local Similarity 58.3%; Pred. No. 2.7e+02;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 QDQMVVQVWP 14
DB 1246 QDRMAPVQNP 1257

RESULT 6
US-09-064-922-2
Sequence 2, Application US/09064922

PATENT NO. 6222095
GENERAL INFORMATION:
APPLICANT: Callis, Judy
APPLICANT: Morley, Cathy K.
TITLE OF INVENTION: Sequences from Auxin-Induced Gene
TITLE OF INVENTION: Products Targeting Fusion Proteins for Degradation
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/064,922
FILING DATE: 22-APR-1998
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:

NAME: Hyman, Laurence J.
REGISTRATION NUMBER: 35,551
REFERENCE/DOCKET NUMBER: 023070-085400US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 156 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-064-922-2

Query Match 46.2%; Score 37; DB 4; Length 156;
Best Local Similarity 58.3%; Pred. No. 23;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 DQDQMVVQVWP 13
DB 112 DQDQMLVQVWP 123

RESULT 7
US-09-064-922-5
Sequence 5, Application US/09064922

PATENT NO. 6222095
GENERAL INFORMATION:
APPLICANT: Callis, Judy
APPLICANT: Morley, Cathy K.
TITLE OF INVENTION: Sequences from Auxin-Induced Gene
TITLE OF INVENTION: Products Targeting Fusion Proteins for Degradation
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/064,922
FILING DATE: 22-APR-1998
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Hyman, Laurence J.
REGISTRATION NUMBER: 35,551
REFERENCE/DOCKET NUMBER: 023070-085400US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 179 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-064-922-5

Query Match 46.2%; Score 37; DB 4; Length 179;
Best Local Similarity 58.3%; Pred. No. 27;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 DQDQMVVQVWP 13
DB 135 DQDQMLVQVWP 146

RESULT 8
US-09-491-362-7
; Sequence 7, Application US/09491362
; Patent No. 6281017
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney B
; APPLICANT: Lange, Bernd M
; TITLE OF INVENTION: 1-DEOXY-D-XYULOSE-5-PHOSPHATE REDUCTOISOMERASE, AND
; TITLE OF INVENTION: METHODS OF USE
; FILE REFERENCE: WS0914977
; CURRENT APPLICATION NUMBER: US/09/491,362
; EARLIER FILING DATE: 2000-01-26
; EARLIER APPLICATION NUMBER: 60/118,349
; EARLIER FILING DATE: 1999-02-03
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 7
; LENGTH: 399
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-491-362-7

Query Match 46.2%; Score 37; DB 4; Length 399;
Best Local Similarity 50.0%; Pred. No. 69;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 DQMYYVQVPW 14
|||1:1:1|
DB 254 QDSSVLAQLGMP 265

RESULT 9
US-08-764-870-16
; Sequence 16, Application US/08764870
; Patent No. 6236946
; GENERAL INFORMATION:
; APPLICANT: Scanlan, Thomas S
; APPLICANT: Baxter, John D
; APPLICANT: Fletterick, Robert J
; APPLICANT: Wagner, Richard L
; APPLICANT: Kushner, Peter J
; APPLICANT: Apriletti, James W
; APPLICANT: West, Brian
; TITLE OF INVENTION: Nuclear Receptor Ligands and Ligand
; TITLE OF INVENTION: Binding Domains
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooley Godward
; STREET: Five Palo Alto Square, 3000 El Camino Real
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,870
; FILING DATE: 13-DEC-1996
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/008,540
; FILING DATE: 13-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/008,543
; FILING DATE: 13-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/008,606

FILING DATE: 14-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Nakamura, Jackie N
REGISTRATION NUMBER: 35,966
REFERENCE/DOCKET INFORMATION:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650)843-5000
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 452 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-764-870-16

Query Match 46.2%; Score 37; DB 4; Length 452;
Best Local Similarity 60.0%; Pred. No. 80;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 DQMYYVQVPW 13
|||1:1:1|
DB 265 DQMAVIOYSW 274

RESULT 10
US-08-980-115-16
; Sequence 16, Application US/08980115
; Patent No. 626622
; GENERAL INFORMATION:
; APPLICANT: Scanlan, Thomas S.
; APPLICANT: Baxter, John D.
; APPLICANT: Fletterick, Robert J.
; APPLICANT: Wagner, Richard L.
; APPLICANT: Kushner, Peter J.
; APPLICANT: Apriletti, James W.
; APPLICANT: West, Brian L.
; APPLICANT: Shiau, Andrew K.
; TITLE OF INVENTION: NUCLEAR RECEPTOR LIGANDS AND LIGAND BINDING DOMAINS
; FILE REFERENCE: UCAL-246/02US
; CURRENT APPLICATION NUMBER: US/08/980,115
; CURRENT FILING DATE: 1997-11-26
; EARLIER APPLICATION NUMBER: 08/764,870
; EARLIER FILING DATE: 1996-12-13
; EARLIER APPLICATION NUMBER: 60/008,606
; EARLIER FILING DATE: 1995-12-14
; EARLIER APPLICATION NUMBER: 60/008,543
; EARLIER FILING DATE: 1995-12-13
; EARLIER APPLICATION NUMBER: 60/008,540
; EARLIER FILING DATE: 1995-12-13
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 16
; LENGTH: 452
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (184)..(437)
; OTHER INFORMATION: minimal ligand binding domain
US-08-980-115-16

Query Match 46.2%; Score 37; DB 4; Length 452;
Best Local Similarity 60.0%; Pred. No. 80;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 DQMYYVQVPW 13
|||1:1:1|
DB 265 DQMAVIOYSW 274

RESULT 11
US-09-491-362-2
; Sequence 2, Application US/09491362
; Patent No. 6281017
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney B
; APPLICANT: Lange, Bernd M
; TITLE OF INVENTION: 1-DEOXY-D-XYLULOSE-5-PHOSPHATE REDUCTOISOMERASE, AND
; TITLE OF INVENTION: METHODS OF USE
; FILE REFERENCE: NSUR14977
; CURRENT APPLICATION NUMBER: US/09/491,362
; CURRENT FILING DATE: 2000-01-26
; EARLIER APPLICATION NUMBER: 60/118,349
; EARLIER FILING DATE: 1999-02-03
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Mentha piperita
US-09-491-362-2

Query Match 46.2%; Score 37; DB 4; Length 475;
Best Local Similarity 50.0%; Pred. No. 84;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 3 ODOMVYGVMP 14
|| | : | : ||
Db 331 QDSSVLAQLGMP 342

RESULT 12
US-09-449-335-2
; Sequence 2, Application US/09449335
; Patent No. 6303365
; GENERAL INFORMATION:
; APPLICANT: Busch, Marco
; APPLICANT: Hain, Rudiger
; APPLICANT: Martin, William
; APPLICANT: Tietjen, Klaus
; TITLE OF INVENTION: Method of determining the activity of
; TITLE OF INVENTION: 1-deoxy-D-xylose-5-phosphate reductoisomerase and
; TITLE OF INVENTION: 1-deoxy-D-xylose-5-phosphate synthase
; FILE REFERENCE: 2020US
; CURRENT APPLICATION NUMBER: US/09/449,335
; CURRENT FILING DATE: 1999-11-24
; EARLIER APPLICATION NUMBER: DE 199 35 967.9
; EARLIER FILING DATE: 1999-07-30
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 477
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-449-335-2

Query Match 46.2%; Score 37; DB 4; Length 477;
Best Local Similarity 50.0%; Pred. No. 85;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 3 ODOMVYGVMP 14
|| | : | : ||
Db 332 QDSSVLAQLGMP 343

RESULT 13
US-09-449-335-6
; Sequence 6, Application US/09449335
; Patent No. 6303365
; GENERAL INFORMATION:

; APPLICANT: Busch, Marco
; APPLICANT: Hain, Rudiger
; APPLICANT: Martin, William
; APPLICANT: Tietjen, Klaus
; TITLE OF INVENTION: Method of determining the activity of
; TITLE OF INVENTION: 1-deoxy-D-xylose-5-phosphate reductoisomerase and
; TITLE OF INVENTION: 1-deoxy-D-xylose-5-phosphate synthase
; FILE REFERENCE: 2020US
; CURRENT APPLICATION NUMBER: US/09/449,335
; CURRENT FILING DATE: 1999-11-24
; EARLIER APPLICATION NUMBER: DE 199 35 967.9
; EARLIER FILING DATE: 1999-07-30
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 477
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-449-335-6

Query Match 46.2%; Score 37; DB 4; Length 477;
Best Local Similarity 50.0%; Pred. No. 85;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 3 ODOMVYGVMP 14
|| | : | : ||
Db 332 QDSSVLAQLGMP 343

RESULT 14
US-09-041-886-11
; Sequence 11, Application US/09041886
; Patent No. 6235872
; GENERAL INFORMATION:
; APPLICANT: Bredesen, Dale E.
; APPLICANT: Rabizadeh, Sharrooz
; TITLE OF INVENTION: Proapoptotic Peptides, Dependence
; TITLE OF INVENTION: Polypeptides and Methods of Use
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESS: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,886
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2626
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 918 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-041-886-11

Query Match 46.2%; Score 37; DB 4; Length 918;
Best Local Similarity 60.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 DQMYVQVWPW 13
| | | | |
DB 731 DQMAVIOYSW 740

RESULT 15

US-09-184-001-4
; Sequence 4, Application US/09184001A
; Patent No. 630333
; GENERAL INFORMATION:
; APPLICANT: MICHALOVICH, DAVID
; APPLICANT: SIMS, MATTHEW A.
; APPLICANT: SHAIKH, NARJIS
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP-30088
; CURRENT APPLICATION NUMBER: US/09/184,001A
; CURRENT FILING DATE: 1998-11-02
; EARLIER APPLICATION NUMBER: 9817479.0
; EARLIER FILING DATE: 1998-08-11
; EARLIER APPLICATION NUMBER: 9806221.9
; EARLIER FILING DATE: 1998-03-23
; SOFTWARE: FASTSEQ for Windows Version 3.0
; NUMBER OF SEQ ID NOS: 4
; SEQ ID NO 4
; LENGTH: 400
; TYPE: PRN
; ORGANISM: HOMO SAPIENS
; FEATURE:
; NAME/KEY:
; LOCATION: (53)(59)(98)
US-09-184-001-4

Query Match 45.0%; Score 36; DB 4; Length 400;
Best Local Similarity 35.7%; Pred. No. 1e+02;
Matches 5; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 YDQDQMYVQVWPW 14
| | | | |
DB 39 WSOGHCTIVKLIWPW 52

Search completed: November 30, 2001, 09:43:39
Job time: 198 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 30, 2001, 23:05:29 ; Search time 3540.92 Seconds
(without alignments)

11726.727 Million cell updates/sec

Title: US-08-846-234-4

Perfect score: 2517
Sequence: 1 AAAAAACACACCTTCTTTA.....TTTGAGAGATTAAAGTGA 2517

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

Genbank1.*
1: gb_ba.*
2: gb_hcg.*
3: gb_in.*
4: gb_cm.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_scs.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_cm.*
20: em_or.*
21: em_ov.*
22: em_pat.*
23: em_ph.*
24: em_pl.*
25: em_ro.*
26: em_scs.*
27: em_sy.*
28: em_un.*
29: em_vl.*
30: em_htgo_hum.*
31: em_htgo_inv.*
32: em_htg_rod.*
33: em_htg_hum.*
34: em_htg_inv.*
35: em_htg_rod.*
36: em_htg_other.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2517	100.0	2517	6	E15707	E15707 Cucumis sat
2	2506	99.6	2569	6	E25430	E25430 Rafifinose s
3	2506	99.6	2569	6	AF073744	AF073744 Cucumis s
4	1041.4	41.4	2690	6	E36417	E36417 Rafifinose s
5	968.6	38.5	2497	6	E24424	E24424 Rafifinose s
6	956.6	38.0	2746	6	E24423	E24423 Rafifinose s
7	886.8	35.2	1762	6	E24425	E24425 Rafifinose s
8	868.4	34.5	1762	6	E36418	E36418 Rafifinose s
9	406.2	16.1	2727	8	PSA311087	AJ131087 Pisum sat
10	392	15.6	3073	8	VANI9024	Y19024 Vigna angul
11	364	14.5	135295	8	AP003282	AP003282 Oryza sat
12	303.8	12.1	928	6	E30796	E30796 Rafifinose s
13	275.8	11.0	2780	6	E25448	E25448 Rafifinose s
14	263.8	10.5	86263	8	AB006702	AB006702 Arabidops
15	252.2	10.0	2669	8	BOPTIN	X79330 B. oleracea
16	237.6	9.4	2673	8	PAM133148	AJ133148 Persica am
17	170.4	6.8	120185	8	AC007138	AC007138 Arabidops
18	170.4	6.8	198220	8	ATCHRIV5	AL161493 Arabidops
19	169.6	6.7	993	6	E24426	E24426 Rafifinose s
20	157.8	6.3	101458	8	F5024	AF296825 Arabidops
21	130.8	5.2	1487	8	CAR271668	AJ271668 Cicer ari
22	111.8	4.4	120868	2	AP003748	AP003748 Oryza sat
23	97	3.9	122095	8	AP002854	AP002854 Oryza sat
24	90.8	3.6	253	6	E25451	E25451 Rafifinose s
25	81.4	3.2	87503	8	ATT8H10	AL133348 Arabidops
26	78.4	3.1	850	8	AF159378	AF159378 Arabidops
27	71.8	2.9	4788	8	BLYSIP1A	M77475 Hordeum vul
28	70.6	2.8	109171	8	AC002328	AC002328 Genomic s
29	66	2.6	90767	2	H0510A06	AL442104 Oryza sat
30	56	2.2	1071	8	CASIMB1P	X95875 C. arletium
31	49.4	2.0	167031	2	AC080013	AC080013 Homo sapi
32	49.4	2.0	170048	2	AC089988	AC089988 Homo sapi
33	47.4	1.9	7218	6	166494	166494 Sequence 14
34	44.8	1.8	9623	1	AE006903	AE006903 Sulfolobu
35	44.6	1.8	140349	9	AC027320	AC027320 Homo sapi
36	42.8	1.7	190349	9	AC009475	AC009475 Homo sapi
37	42.8	1.7	191904	2	AC009956	AC009956 Homo sapi
38	42.4	1.7	2239	3	DDU07817	U07817 Dictyostel
39	42.4	1.7	179862	9	CNS05TEE	AL358913 Human chr
40	42	1.7	165916	9	AC008901	AC008901 Homo sapi
41	42	1.7	191652	9	AL354668	AL354668 Human DNA
42	41.8	1.7	160903	2	AP002776	AP002776 Homo sapi
43	41.6	1.7	40931	2	AC027349	AC027349 Homo sapi
44	41.6	1.7	133387	2	AP004061	AP004061 Oryza sat
45	41.6	1.7	155774	2	AL592202	AL592202 Dantio rer

ALIGNMENTS

RESULT	1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
E15707	E15707	Cucumis sativus mRNA for raffinose synthase, complete cds.	2517 bp	E15707	E15707.1	GI:5710390	JP 1998084973-A/1.	unclassified.	1 (bases 1 to 2517)	Oosumi,C., Nozaki,T. and Kida,T.	RAFFINOSE SYNTHASE GENE, PRODUCTION OF RAFFINOSE AND TRANSFORMED	ATINOMOTO CO INC	OS Cucumis sativus (cucumber) PN JP 1998084973-A/1 PD 07-APR-1998 PF 28-APR-1997 JP 1997111124 PI 26-APR-1996 JP 96P 107682, 26-JUL-1996 JP 96P 198079

OOSUMI CHIEKO, NOZAKI JINJI, KIDA TAKAO
PC C12N15/09,A01H5/00,C12N9/00;
CC strandedness: Double;
CC topology: Linear;
FH key Location/Qualifiers
FT source 1..2517
FT /organism='Cucumis sativus'
FT /flsue_type='leaf'
FT CDS 56..2410
FT /product='raffinose synthase'.
FEATURES
source 1..2517
Location/Qualifiers
BASE COUNT 613 a 524 c 642 g 738 t
ORIGIN
Query Match 100.0%; Score 2517; DB 6; Length 2517;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2517; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAAAAACACCCCTCTTTAGTTTGGGTTTGGTTTCTTCTTCTTCTCACAAATGGC 60
DB 1 AAAAAACACCCCTCTTTAGTTTGGGTTTGGTTTCTTCTTCTTCTCACAAATGGC 60
QY 61 TCTTACTTTAAAAATGTTGGCTCCAAACGATGTTCAATTGATGGCTTAATGACATGC 120
DB 61 TCTTACTTTAAAAATGTTGGCTCCAAACGATGTTCAATTGATGGCTTAATGACATGC 120
QY 121 GTACAGCTTGCATCGACGATCGATTCAGTGTGAACGTCATTCGTTCTGTCGA 180
DB 121 GTACAGCTTGCATCGACGATCGATTCAGTGTGAACGTCATTCGTTCTGTCGA 180
QY 181 TGTTCCTGAGAACATTGTTCTCTCTCTCCGTACATTCGATACACAAGTCCCGGT 240
DB 181 TGTTCCTGAGAACATTGTTCTCTCTCTCCGTACATTCGATACACAAGTCCCGGT 240
QY 241 TTGGGTTGTTGCTTTGTTGGATTGCGACGCGTGGAGACGATAGCCGACATGTTGTTTC 300
DB 241 TTGGGTTGTTGCTTTGTTGGATTGCGACGCGTGGAGACGATAGCCGACATGTTGTTTC 300
QY 301 GATTGGAACTGAAAGATATTCGTTTATGAGTATTTTTCAGTTTAAAGTTTGGTGGAC 360
DB 301 GATTGGAACTGAAAGATATTCGTTTATGAGTATTTTTCAGTTTAAAGTTTGGTGGAC 360
QY 361 TACACACTGGTGTGGTGAATGTTGGGATCTTGAATCGAGACTAGATTTGATCCT 420
DB 361 TACACACTGGTGTGGTGAATGTTGGGATCTTGAATCGAGACTAGATTTGATCCT 420
QY 421 TGAAGATCGATCTGTTGAGCGATGATGTTTCTCTCTCCGATGTTGTCGAGAGTGGTTC 480
DB 421 TGAAGATCGATCTGTTGAGCGATGATGTTTCTCTCTCCGATGTTGTCGAGAGTGGTTC 480
QY 481 CCGAACCTGATTCAGCCTGGGATGATGACTTGTGATGTTTGTGTGAGAGTGGTTC 540
DB 481 CCGAACCTGATTCAGCCTGGGATGATGACTTGTGATGTTTGTGTGAGAGTGGTTC 540
QY 541 GTCCAAAGTTGTTGATGATCGTTCGAAAGTATGTTTATCTGATCTGATGATGATCC 600
DB 541 GTCCAAAGTTGTTGATGATCGTTCGAAAGTATGTTTATCTGATCTGATGATGATCC 600
QY 601 GTTTGGCACTGTTAAAGAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
DB 601 GTTTGGCACTGTTAAAGAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
QY 661 GTTGGAGAGAGAGATCCACAGGATATCTGGACAAATTCGTTGGTGTGACAGTGGAGGC 720
DB 661 GTTGGAGAGAGAGATCCACAGGATATCTGGACAAATTCGTTGGTGTGACAGTGGAGGC 720
QY 721 GTTTTACCTAACGTTATCTCACAGGCGTATATGAAGGCGTGAAGCATCTGTCGACGG 780
DB 721 GTTTTACCTAACGTTATCTCACAGGCGTATATGAAGGCGTGAAGCATCTGTCGACGG 780

DB 721 GTTTTACCTAACGTTATCTCACAGGCGTATATGAAGGCGTGAAGCATCTGTCGACGG 780
QY 781 CGGTGGCTCCCGGTTTATGTTTCTTAATCGACGATGTTGGCAATTCATCGACAGATTC 840
DB 781 CGGTGGCTCCCGGTTTATGTTTCTTAATCGACGATGTTGGCAATTCATCGACAGATTC 840
QY 841 GGATCCCATCACAAGAGAGATGAACCAACCGTCCGCGGAGCAATGCTTCGCC 900
DB 841 GGATCCCATCACAAGAGAGATGAACCAACCGTCCGCGGAGCAATGCTTCGCC 900
QY 901 TCTTTTGAATTCGAGAGATGAACCAATTCGTTGATGATGATGATGATGATGATGATGAT 960
DB 901 TCTTTTGAATTCGAGAGATGAACCAATTCGTTGATGATGATGATGATGATGATGATGAT 960
QY 961 CCCCCGAGCCGCGCAGAGAGAGATGAAGCGTTTATAGATGAATCAAGAGAGTTTAA 1020
DB 961 CCCCCGAGCCGCGCAGAGAGAGATGAAGCGTTTATAGATGAATCAAGAGAGTTTAA 1020
QY 1021 GACTGTGAGCATGTTTATGTTTGGCATGCTTTGTGTGATATTTGGGTTGGCTTGGCCC 1080
DB 1021 GACTGTGAGCATGTTTATGTTTGGCATGCTTTGTGTGATATTTGGGTTGGCTTGGCCC 1080
QY 1081 GCAGTGGCTTGGCTTGGCTGAGGACGTTGATTCAGCCAGTGTCTTACAGAGGCTGCA 1140
DB 1081 GCAGTGGCTTGGCTTGGCTGAGGACGTTGATTCAGCCAGTGTCTTACAGAGGCTGCA 1140
QY 1141 GATGACGATGAGAGATTTGGCGGTGATTAAGATGTTCTTATTAAGTGGGCTGTGCC 1200
DB 1141 GATGACGATGAGAGATTTGGCGGTGATTAAGATGTTCTTATTAAGTGGGCTGTGCC 1200
QY 1201 GCCGGAGAGGCTGAGAGATGTCACAGACATTCATGCTCATTTGGAAAAAGTTGGAT 1260
DB 1201 GCCGGAGAGGCTGAGAGATGTCACAGACATTCATGCTCATTTGGAAAAAGTTGGAT 1260
QY 1261 CGAGGCTTAAAGTATGACCTTATCCACTATTGAGAGATTTGTGMAAGCATATGAGG 1320
DB 1261 CGAGGCTTAAAGTATGACCTTATCCACTATTGAGAGATTTGTGMAAGCATATGAGG 1320
QY 1321 GAGAGTGATTTGGCAAAAGCATATTAACAAGCAATGACCAATCAATTAATAACATTT 1380
DB 1321 GAGAGTGATTTGGCAAAAGCATATTAACAAGCAATGACCAATCAATTAATAACATTT 1380
QY 1381 TAAAGAAATGAGATCAATTCAGATGAGAAACATTTGAAGACTTCATGTTCTTGGCAC 1440
DB 1381 TAAAGAAATGAGATCAATTCAGATGAGAAACATTTGAAGACTTCATGTTCTTGGCAC 1440
QY 1441 GGAAGCTATCTCTGTTGGTGTGATGACTTTTGTGTCACGAGACCCCTGTGTA 1500
DB 1441 GGAAGCTATCTCTGTTGGTGTGATGACTTTTGTGTCACGAGACCCCTGTGTA 1500
QY 1501 TCCAAACGTAAGTTTGGCTCCAGAGATGTCACATGTTTATGTTGTCACAGACACTT 1560
DB 1501 TCCAAACGTAAGTTTGGCTCCAGAGATGTCACATGTTTATGTTGTCACAGACACTT 1560
QY 1561 GTGATGGGAACTTCATCCACCTGAGCTGGGATATGTTCCAAATCCACCAACCTTGTGC 1620
DB 1561 GTGATGGGAACTTCATCCACCTGAGCTGGGATATGTTCCAAATCCACCAACCTTGTGC 1620
QY 1621 CGCCTTCATGCTGCTCTGAGCCTATCTGTGTGCGCCGATCATGTATGTATGATCTGT 1680
DB 1621 CGCCTTCATGCTGCTCTGAGCCTATCTGTGTGCGCCGATCATGTATGTATGATCTGT 1680
QY 1681 GGAAGACATATCTTGATCTTGAAAAAACTGATGCTTCTTATGATGATGATGATGATGAT 1740
DB 1681 GGAAGACATATCTTGATCTTGAAAAAACTGATGCTTCTTATGATGATGATGATGATGAT 1740
QY 1741 AAGTGAAGTATGACCTCCGACTCGGATGTTGTTGTAAGACCTTTCATATAGG 1800
DB 1741 AAGTGAAGTATGACCTCCGACTCGGATGTTGTTGTAAGACCTTTCATATAGG 1800
QY 1801 AGAAGTATGCTTAAAGATTTGGAATCTCAACAGTTCACTGAGATGATGATGATGATGAT 1860
DB 1801 AGAAGTATGCTTAAAGATTTGGAATCTCAACAGTTCACTGAGATGATGATGATGATGAT 1860

Oy	1861	CTGCCAAGGAGAGATGGTGTCTGTGTGACACAGCCCGCCAACAATGCTTTTCACAAATACTC	1920
Db	1861	CTGCCAAGGAGAGATGGTGTCTGTGTGACACAGCCCGCCAACAATGCTTTTCACAAATACTC	1920
Oy	1921	AAAAGAGTGACATCCAAAACTAACCCAAAAGACATAGATGGCACAGTGGAGAAAAACC	1980
Db	1921	AAAAGAGTGACATCCAAAACTAACCCAAAAGACATAGATGGCACAGTGGAGAAAAACC	1980
Oy	1981	TATCTCTATTGAAGCGTTTAAAAACCTTTGCCTTTACTCTATCAAGCCAAAAACTTAT	2040
Db	1981	TATCTCTATTGAAGCGTTTAAAAACCTTTGCCTTTACTCTATCAAGCCAAAAACTTAT	2040
Oy	2041	CCTCTCCAAAGCCCTCTCAAGAATCTTGACATAGCTCTTTGACCCCATTTGGAATTCGAGCTCAT	2100
Db	2041	CCTCTCCAAAGCCCTCTCAAGAATCTTGACATAGCTCTTTGACCCCATTTGGAATTCGAGCTCAT	2100
Oy	2101	CACCTTTTCACCAAGTAGACCAAACATCATCCAACTTCTCTACACTTTGGCCCCCAATTGGGCT	2160
Db	2101	CACCTTTTCACCAAGTAGACCAAACATCATCCAACTTCTCTACACTTTGGCCCCCAATTGGGCT	2160
Oy	2161	GGTGAACATGCTTTAACTAGTGGAGCCATCCAACTGTGTGAGACTATGACGATGACCTTAAG	2220
Db	2161	GGTGAACATGCTTTAACTAGTGGAGCCATCCAACTGTGTGAGACTATGACGATGACCTTAAG	2220
Oy	2221	CTCACTCGAGATTTGGGTGTCAAAAGGGGTGTGGTGTGATGGCGATTTTGGCATCGAAAAAAC	2280
Db	2221	CTCACTCGAGATTTGGGTGTCAAAAGGGGTGTGGTGTGATGGCGATTTTGGCATCGAAAAAAC	2280
Oy	2281	AAGGCGTTGCTCTAATVTGATGGGAGAGATTTGGGTTCAAGTATGATCAGAGCAACCAATGT	2340
Db	2281	AAGGCGTTGCTCTAATVTGATGGGAGAGATTTGGGTTCAAGTATGATCAGAGCAACCAATGT	2340
Oy	2341	GGTGGTTCAAGTGGCCATGGCCCAATTGATTTCTTCATCGGTGGCATTTTCGGTTATCCAGTA	2400
Db	2341	GGTGGTTCAAGTGGCCATGGCCCAATTGATTTCTTCATCGGTGGCATTTTCGGTTATCCAGTA	2400
Oy	2401	CTTGTTTTTAATTTTTAATTAATGTAAGCTCAATGATTTGTTGGTGGCGTGTATGTTGCT	2460
Db	2401	CTTGTTTTTAATTTTTAATTAATGTAAGCTCAATGATTTGTTGGTGGCGTGTATGTTGCT	2460
Oy	2461	ATCAATGTATTTCTCTCCAAAAAATAATGTATGTATTTGGAGAGTAATTAAGTGA	2517
Db	2461	ATCAATGTATTTCTCTCCAAAAAATAATGTATGTATTTGGAGAGTAATTAAGTGA	2517
RESULT	2		
E25430			
LOCUS	E25430	2569 bp	DNA
DEFINITION	Raffinose synthase gene, process for producing raffinose and transformed plant.	PAT	07-FEB-2001
ACCESSION	E25430		
VERSION	E25430.1	GI:13024726	
KEYWORDS	JP 199123080-A/I.		
SOURCE	unidentified.		
ORGANISM	unidentified.		
REFERENCE	1 (bases 1 to 2569)		
AUTHORS	Chleko,O.J.N.N. and Kida.		
TITLE	Raffinose synthase gene, process for producing raffinose and transformed plant		
JOURNAL	Patent: JP 199123080-A 1 11-MAY-1999;		
COMMENT	AJINOMOTO CO INC		
	OS Cucumis sativus (Cucumber)		
	PN JP 1999123080-A/I		
	PD 11-MAY-1999		
	PR 24-OCT-1997 JP 1997292969		
	PI CHIEKO OZUMI,AJIHISU NOZAKI,TAKAO KIDA		
	PC C12N15/09,A01H5/00,C12N9/00//((C12N9/00.C12N1.19).C12N15/00 CC		
	Strandedness: Double:		
	CC Topology: Linear:		
	FH key Location/Qualifiers		

FEATURES	FT	CDS	Location/Qualifiers
SOURCE		1..2569	/organism="unidentified" /db_xref="taxon:32644"
BASE COUNT	641 a	527 c	643 g 755 t 3 others
ORIGIN			
Query Match 99.6%; Score 2506; DB 6; Length 2569;			
Best Local Similarity 100.0%; Pred. No. 0;			
Matches 2517; Conservative 0; Mismatches 0; Indels 1; Gaps 1;			
OY	1	AAAAAACACCCCTTCTTTAGTTTTGGGTTTGTTCCTCTTCTCACAATGGC	60
Db	1	AAAAAAAAACCCTCTTTAGTTTTTGGGTTTGTTCCTCTTCTCACAATGGC	60
OY	61	TCCCTGATTAAAATGAGGCGCTCCACAGTAGTTTCATTGATGGCTTAAATGACATGTC	120
Db	61	TCCCTGATTAAAATGAGGCGCTCCACAGTAGTTTCATTGATGGCTTAAATGACATGTC	120
OY	121	GTCACCGTTTGCATTCAGCGAGTCCGATTTCACTGTGAACGCTCATTCGTTCTGCCGA	180
Db	121	GTCACCGTTTGCATTCAGCGAGTCCGATTTCACTGTGAACGCTCATTCGTTCTGCCGA	180
OY	181	TGTTCTCTAGAACAATGTTGCTTTCCTTCCTTCCTCCGACCTTCGATAGACAATGCCCGGT	240
Db	181	TGTTCTCTAGAACAATGTTGCTTTCCTTCCTTCCTCCGACCTTCGATAGACAATGCCCGGT	240
OY	241	TTTCGGTTGGTTCCTTTGTTGGAATTTGACGCGCTCGGACACTGATPAGCCGACATGTTGTTTC	300
Db	241	TTTCGGTTGGTTCCTTTGTTGGAATTTGACGCGCTCGGACACTGATPAGCCGACATGTTGTTTC	300
OY	301	GATTGGGAAGCTGAANGATATTCCGTTTAATGATATTTTCAGGTTTAAAGTTGGTGAC	360
Db	301	GATTGGGAAGCTGAANGATATTCCGTTTAATGATATTTTCAGGTTTAAAGTTGGTGAC	360
OY	361	TACACACGCGGTGGTCGCAAAATGAGGGGATCTTGAATCGGAGACTCGAATGTGATCCT	420
Db	361	TACACACGCGGTGGTCGCAAAATGAGGGGATCTTGAATCGGAGACTCGAATGTGATCCT	420
OY	421	TGAGAAGTCAGATTTGTGTCGACCGTATGTTTCCCTCTTCGATCGTTGAGGAGCCGTT	480
Db	421	TGAGAAGTCAGATTTGTGTCGACCGTATGTTTCCCTCTTCGATCGTTGAGGAGCCGTT	480
OY	481	CCGAACCTTCGATTCACGCTCGGGAGATGACTTTGTCCATGTTTGTGTGACAGTGGTTC	540
Db	481	CCGAACCTTCGATTCACGCTCGGGAGATGACTTTGTCCATGTTTGTGTGACAGTGGTTC	540
OY	541	GTCGAAGGTTGTGATGCATCGTCCGAAGTATGTTGTCATGTCGAGTGGATGATCC	600
Db	541	GTCGAAGGTTGTGATGCATCGTCCGAAGTATGTTGTCATGTCGAGTGGATGATCC	600
OY	601	GTTTGCACATTGTTAAAGAGCGGATGAAGATCGTAGAGACCCATCTTGGAACTTTTCGCTT	660
Db	601	GTTTGCACATTGTTAAAGAGCGGATGAAGATCGTAGAGACCCATCTTGGAACTTTTCGCTT	660
OY	661	GTTTGGAGAGAAGACTCCACCAAGGATATGTGGACAAATTCGGTTGGTGACCTGGGACGC	720
Db	661	GTTTGGAGAGAAGACTCCACCAAGGATATGTGGACAAATTCGGTTGGTGACCTGGGACGC	720
OY	721	GTTTTCACGTAACGATGTTCAATCCACAGGCGCTAATPAGAAGCGTAGAGCATCTGTGACGG	780
Db	721	GTTTTCACGTAACGATGTTCAATCCACAGGCGCTAATPAGAAGCGTAGAGCATCTGTGACGG	780
OY	781	CGGTTGTCTCCCGGTTTATGTCCTAATCGACGATGTTGGCAATCCATTCGACACAGATTTC	840
Db	781	CGGTTGTCTCCCGGTTTATGTCCTAATCGACGATGTTGGCAATCCATTCGACACAGATTTC	840
OY	841	GGATCCCATCACCAAAGAAGATATAACCAACGCTCGCGCGGACCAATGCTCCCTCGCG	900
Db	841	GGATCCCATCACCAAAGAAGATATAACCAACGCTCGCGCGGACCAATGCTCCCTCGCG	900

```
QY 901 TCTTTGAAATTCAGAGAAATTCACAAATTCGGTACCTACCTCATCCCAAGGCCACCGG 960
|||||
Db 901 TCTTTGAAATTCAGAGAAATTCACAAATTCGGTACCTACCTCATCCCAAGGCCACCGG 960
QY 961 CCCCCGCGCGGAGAGAGGAGATGAAGGGCTTATAGATGAATCTCAAGAGAGATTAA 1020
|||||
Db 961 CCCCCGCGCGGAGAGAGGAGATGAAGGGCTTATAGATGAATCTCAAGAGAGATTAA 1020
QY 1021 GACTGTGACATGTTTATGTTTGGCATGCTTGTGTGATATTGGGGTGGCTTCCGCC 1080
|||||
Db 1021 GACTGTGACATGTTTATGTTTGGCATGCTTGTGTGATATTGGGGTGGCTTCCGCC 1080
QY 1081 GCAGGTGCTGGCTTGGCTTGAAGCAGCTGTGATTCACCCAGTCTTCCACAGGGCTGCA 1140
|||||
Db 1081 GCAGGTGCTGGCTTGGCTTGAAGCAGCTGTGATTCACCCAGTCTTCCACAGGGCTGCA 1140
QY 1141 GATGACATGAGAGATTTGGCGGTGATTAAGATTGTTCTCATTAAGTGGGGGTGGTCCC 1200
|||||
Db 1141 GATGACATGAGAGATTTGGCGGTGATTAAGATTGTTCTCATTAAGTGGGGGTGGTCCC 1200
QY 1201 GCCGGAAGAGCTGAGAGATGTACGAAGGACTTCATGCTCATTTGGAAAAAGTTGGGAT 1260
|||||
Db 1201 GCCGGAAGAGCTGAGAGATGTACGAAGGACTTCATGCTCATTTGGAAAAAGTTGGGAT 1260
QY 1261 CGACGGTGTAAAGATTGACCTTATCCACTATTGAGATGTTGTGTAAGACTATGAGG 1320
|||||
Db 1261 CGACGGTGTAAAGATTGACCTTATCCACTATTGAGATGTTGTGTAAGACTATGAGG 1320
QY 1321 GAGAGTGGATTTGGCAAAAGCATATTACAAAGCAATGACCAATCAATTAATTAACATT 1380
|||||
Db 1321 GAGAGTGGATTTGGCAAAAGCATATTACAAAGCAATGACCAATCAATTAATTAACATT 1380
QY 1381 TAAAGAAATGAGATGCATTCGAATGAGAAATGTAAAGACTTCATGTTCTTGGGAC 1440
|||||
Db 1381 TAAAGAAATGAGATGCATTCGAATGAGAAATGTAAAGACTTCATGTTCTTGGGAC 1440
QY 1441 GGAAGCTATCTCTCTGGTGTGTGGTGTGATGACTTTGGTGACGAGCCCTCTGGTGA 1500
|||||
Db 1441 GGAAGCTATCTCTCTGGTGTGTGGTGTGATGACTTTGGTGACGAGCCCTCTGGTGA 1500
QY 1501 TCCAAAGGATGCTTGGTGGCTCCGAAGGATGTACATGTTATTTGGTGGCAAGCAGCT 1560
|||||
Db 1501 TCCAAAGGATGCTTGGTGGCTCCGAAGGATGTACATGTTATTTGGTGGCAAGCAGCT 1560
QY 1561 GTGGATGGGAACTTCATCCACCCTGACTGGGATATGTTCCAAATCCACCACCCTTGTGC 1620
|||||
Db 1561 GTGGATGGGAACTTCATCCACCCTGACTGGGATATGTTCCAAATCCACCACCCTTGTGC 1620
QY 1621 CGCCTTCATGCTGCTCTGTCAGCCATCTGTGTGGGCCCATCTATGTTAGTATCTGT 1680
|||||
Db 1621 CGCCTTCATGCTGCTCTGTCAGCCATCTGTGTGGGCCCATCTATGTTAGTATCTGT 1680
QY 1681 GGGAAAGATTAACCTTGTGTCCTTGAAAAAACTAGTCTCTCTGATGATGATCCTTCG 1740
|||||
Db 1681 GGGAAAGATTAACCTTGTGTCCTTGAAAAAACTAGTCTCTCTGATGATGATCCTTCG 1740
QY 1741 AAGTGAAGTATGACATCCCGACTCCGACTCCGATTTGTTTGAAGACCCCTTGCATATG 1800
|||||
Db 1741 AAGTGAAGTATGACATCCCGACTCCGACTCCGATTTGTTTGAAGACCCCTTGCATATG 1800
QY 1801 AGAAACTATGCTTAAAGATTGGAATCTCACAAGTTCACCTGAGAGTATGGTGCATTCAA 1860
|||||
Db 1801 AGAAACTATGCTTAAAGATTGGAATCTCACAAGTTCACCTGAGAGTATGGTGCATTCAA 1860
QY 1861 CTGCAAGAGAGAGATGCTGTCGAGACACGCGCAACCAATGCTTTCACATATCTC 1920
|||||
Db 1861 CTGCAAGAGAGAGATGCTGTCGAGACACGCGCAACCAATGCTTTCACATATCTC 1920
QY 1921 AAAAGAGTGAATCCAAATTAACCCAAAAGACATAGATGGCAGCTGGAGAAAACCC 1980
|||||
Db 1921 AAAAGAGTGAATCCAAATTAACCCAAAAGACATAGATGGCAGCTGGAGAAAACCC 1980
QY 1981 TATCTATATGAAGCGTTAAAAACCTTTGGCTTTACTCTATATCAAGCCAAAAAATTA 2040
|||||
```

```
Db 1981 TATCTATATGAAGCGTTAAAAACCTTTGGCTTTACTCTATATCAAGCCAAAAAATTA 2040
|||||
QY 2041 CCTCTCAAGCCCTCTCAAGATCTTGACATAGCTCTTGACCATTCGAATTCGAGCTCAT 2100
|||||
Db 2041 CCTCTCAAGCCCTCTCAAGATCTTGACATAGCTCTTGACCATTCGAATTCGAGCTCAT 2100
QY 2101 CACTGTTTCACACAGTGCACAAACATCCAAACTCTCTACACTTGGCCCAATGGGCT 2160
|||||
Db 2101 CACTGTTTCACACAGTGCACAAACATCCAAACTCTCTACACTTGGCCCAATGGGCT 2160
QY 2161 GGTAACATGCTTAACACTAGTGGAGCCATCCAAATCTGTGACTATGACATGACCTAAG 2220
|||||
Db 2161 GGTAACATGCTTAACACTAGTGGAGCCATCCAAATCTGTGACTATGACATGACCTAAG 2220
QY 2221 CTCAGTGAGATTGGTGTCAAAAGGGTGTGTGAGATCCGATATTTCATGCAAAAAACC 2280
|||||
Db 2221 CTCAGTGAGATTGGTGTCAAAAGGGTGTGTGAGATCCGATATTTCATGCAAAAAACC 2280
QY 2281 AAGGCTTGTGCTATTATGATGGGAGATGTTGGTTCAGATATGATCAGAACCAATGGT 2340
|||||
Db 2281 AAGGCTTGTGCTATTATGATGGGAGATGTTGGTTCAGATATGATCAGAACCAATGGT 2340
QY 2341 GGTTGTTCAAGTGGCCATGGCCAAATTTGATCTCATCGGGTGGCATTTGGTTATCGAGTA 2400
|||||
Db 2341 GGTTGTTCAAGTGGCCATGGCCAAATTTGATCTCATCGGGTGGCATTTGGTTATCGAGTA 2400
QY 2401 CTGTGTTTAAATTTTAAATTAATGTA-AGCTCAATGATTGTTGTTGCTGCTGTTTGC 2459
|||||
Db 2401 CTGTGTTTAAATTTTAAATTAATGTAATGTAAGCTCAATGATTGTTGTTGCTGCTGTTTGC 2460
QY 2460 TATCAATGATATTCTCTCCAAAAAAAATTTATGTTATTTGGAGAGATTAATTAAGTA 2517
|||||
Db 2461 TATCAATGATATTCTCTCCAAAAAAAATTTATGTTATTTGGAGAGATTAATTAAGTA 2518
|||||

RESULT 3
AF073744 2569 bp mRNA PLN 05-JAN-1999
LOCUS AF073744
DEFINITION Cucumis sativus raffinose synthase (Rfs) mRNA, complete cds.
ACCESSION AF073744
VERSION AF073744.1 GI:4106394
KEYWORDS cucumber.
SOURCE Cucumis sativus
ORGANISM Cucumis sativus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotis;
Rosidae; eurosids I; Cucurbitales; Cucurbitaceae; Cucumis.
REFERENCE 1 (bases 1 to 2569)
AUTHORS Ohsuni,C., Nozaki,J. and Kida,T.
TITLE Direct Submission
JOURNAL Submitted (22-JUN-1998) Central Research Laboratories, Ajinomoto
Co., Inc., 1-1, Suzuki-cho, Kawasaki-ku, Kawasaki 210-8681, Japan
FEATURES
source
1..2569
/organism="Cucumis sativus"
/db_xref="taxon:3659"
/lssue_type="leaf"
/dev_stage="mature"
1..2569
/gene="Rfs"
56..2410
/EC_number="2.4.1.82"
/codon_start=1
/product="raffinose synthase"
/protein_id="AAD02832.1"
/db_xref="GI:4106395"
/translacion="MAPSPKNGSGSNVSPGLNDMSFPAIDGSDFTVNGHSPFLSDVP
ENIVASPSPTSIDKSPVSGCEVGPASEPDSRHVSIKLDKIRFMSIFRRKVMY
THHWKVGSDLESFTQIVILEKSDSGRPVFLPIYBGPRTSIQPDDDFVDCVCS
GSSKVVNASFESMLYLHAQDDPFLALVEAKIYVTHLGFRLLEKTPGIVKFGVC
TWDAFYLTIVHOGVIEGVRLYVDGCPPLVLIDDMQSIGHDSDPITKEGMQTVAG
```


Db	1921	AAAACGAGTGCACATCCAAACTAAACCCAAAGACATGAAATGGCAGAGTGGAGAAAACCC	1980
Qy	1981	TATCTCTATTGGAAGCGCTTAAACCTTTGGCGTTTACCCTATATCAAGCCAAAACCTTAT	2040
Db	1981	TATCTCTATTGGAAGCGCTTAAACCTTTGGCGTTTACCCTATATCAAGCCAAAACCTTAT	2040
Qy	2041	CCCTCTCCAAAGCCCTCTCAAGATCTTGACATAGCTCTTGACCATTGCAATTCGAGCTCAT	2100
Db	2041	CCCTCTCCAAAGCCCTCTCAAGATCTTGACATAGCTCTTGACCATTGCAATTCGAGCTCAT	2100
Qy	2101	CACGTGTTACACAGTGCACAAACTCATCAAACTTCTCTACACTTTGGCCCAATTTGGGCT	2160
Db	2101	CACGTGTTACACAGTGCACAAACTCATCAAACTTCTCTACACTTTGGCCCAATTTGGGCT	2160
Qy	2161	GGTGAACATGCTTAAACACTAGTGGAGCCATCCAAATCTGTGACATATGACATGACCTTAA	2220
Db	2161	GGTGAACATGCTTAAACACTAGTGGAGCCATCCAAATCTGTGACATATGACATGACCTTAA	2220
Qy	2221	CTCAGTGCAGATTTGGTGTCAAAGGGTGGGAGATGGAGATATTTGGCATGGAACCAACC	2280
Db	2221	CTCAGTGCAGATTTGGTGTCAAAGGGTGGGAGATGGAGATATTTGGCATGGAACCAACC	2280
Qy	2281	AAGGCGCTTGTCGATTTGATATGGAGGAGATGTTGGGTTCAAGTATGATCAGACCAATGGT	2340
Db	2281	AAGGCGCTTGTCGATTTGATATGGAGGAGATGTTGGGTTCAAGTATGATCAGACCAATGGT	2340
Qy	2341	GGTGGTTCAGAGTGCACATGCGCCAAATGATCTTCATCGGGTGGCATTTGGGTTATTCGAGTA	2400
Db	2341	GGTGGTTCAGAGTGCACATGCGCCAAATGATCTTCATCGGGTGGCATTTGGGTTATTCGAGTA	2400
Qy	2401	CTTGTTTTTAATTTTTATTTATATGTA - AGCTCAATGATTTGTTGTTGCGCTGTTGTTGC	2459
Db	2401	CTTGTTTTTAATTTTTATTTATATGTA - AGCTCAATGATTTGTTGTTGCGCTGTTGTTGC	2459
Qy	2460	TATCAATGATATTTCTCTCCAAAGAAATTAATGTAATTTGGAGAGTAAATTAAGTGA	2517
Db	2461	TATCAATGATATTTCTCTCCAAAGAAATTAATGTAATTTGGAGAGTAAATTAAGTGA	2518
RESULT	4		
LOCUS	E36417	2690 bp	DNA
DEFINITION	E36417	Raffinose synthase gene.	
ACCESSION	E36417		
VERSION	E36417.1	GI:13019215	
KEYWORDS	JP 2000014389-A/1.		
SOURCE	unidentified.		
ORGANISM	unclassified.		
REFERENCE	1. (bases 1 to 2690)		
AUTHORS	Eljjiro,W.K.O.O.		
TITLE	Raffinose synthase gene		
JOURNAL	Patent: JP 2000014389-A 1 18-JAN-2000;		
COMMENT	SUMITOMO CHEM CO LTD		
	OS Brassica juncea		
	PN JP 2000014389-A/1		
	PD 18-JAN-2000		
	PF 10-DEC-1998 JP 1998351246		
	PR		
	PI EJIRNO WATANABE,KENJI OEDA		
	PC C12N15/09,A01H5/00,C12N1/21,C12N5/10,C12N9/88,C12Q1/68,PC		
	(C12N1/21,C12R1:19),(C12N5/10,C12R1:91),(C12N9/88,C12R1:19),PC		
	(C12N1/21,C12R1:91),C12N15/00,C12N5/00,C12N5/00,C12R1:91) CC		
	FH key	Location/Qualifiers	
	FT CDS	Location/Qualifiers (134)..(2467).	
FEATURES	source	Location/Qualifiers	
	1..2690	/organism="unidentified"	
	/db_xref="taxon:32644"		
BASE COUNT	648 a 608 c 668 g 765 t	1 others	
ORIGIN			

Query Match	41.4%: Score 1041.4: DB 6: Length 2690;
Best Local Similarity	68.3%: Pred. No. 7.3e-291:
Matches 1530; Conservative	0; Mismatches 676; Indels 33; Gaps 5;
QY 135 TCAGAGCATCGGATTCACATGTGAGACGGGTCAATTCGTTTCGTCCGATGTTCTGTAGACACA 194	
DB 210 TAGAGGTTCCGATCTCCTAGACAAATGGTCTGATTAACGATGTACCGGTTAAGC 269	
QY 195 TTGTTGCTTCTCCCTTCGCTAC-----ACTTCGATAGACAAAGTCCCGGTTTGG 244	
DB 270 TGACGTCTACCTGCTTACCTTACCTAGCTGACAAAGACGGAGAACCGGTTGAGCGCTCG 329	
QY 245 GTTGTGTTGCTTTGTGATTCGACCGCTCG--GAACGTATAGCCGATGTTTTCGA 302	
DB 330 CTGGTTTCATTTCATCGGGTTTAATCTCGACGGGTGAGCCACGAAGCCGACATGTTTTCGA 389	
QY 303 TTGGGAAGCTGGAAGGATATTCGTTATGAGATATTTTCAGGTTTAAAGTTTGGTGGACTA 362	
DB 390 TCGGTAACTCAGGGGTATTCGATTTCATGACGATATTCGTTTCGTAAGTTTGGTGGACTA 449	
QY 423 AGAAGTCAGATTCGTGTGACCGTATGTTTTCCTTCTTCGATTCGTTGAGGACCGCTCC 482	
DB 510 AGAAGCTCCGGGTGCGGGTCTCCTTATGTTCTTCTTCGCGCTTCTTGAAGGCTCTTTC 569	
QY 483 GAACCTCGATTCAGCCTGGGGATGATGACTTGTTCGATGTTTGTCTGAGAGTGGTTCGT 542	
DB 570 GTTCATCTCTTTCAGCCTGGGGAGACGATGACGTGGCGGTTTGTGTGAATCCGGGGTGA 629	
QY 543 CGAAAGTTTGATGATCATGTTCCGAAATATGTGTATCTTCATGCTGTGATGATCCGT 602	
DB 630 CCCAGGTGACGGGTGCGAGTTTCGTCAAGTTGTGTATGTTTCACGCCGAGACATCCGT 689	
QY 603 TTGCACCTGTTTAAAGAGCGGATGAAGATCTGTAGAGACCCATCTTGGAACCTTTTCGTTGT 662	
DB 690 TCAAGCTCTCGGAAGACGCGATGAAGGTGGTTAAGGTTTCATATAMACACCTTCAAGCTCT 749	
QY 663 TGGAGGAGAAGATCCTCAGGATATCTGGAACAAATTTGGTTGGTGGACGCGGAGCGCT 722	
DB 750 TGGAGAAGGAAGACRCCGCGGAATCTGTGATTAATTTGGGTGGTGTGACGCTGGGATTCGT 809	
QY 723 TTTTACCTTAAGGTTTCTCCACAGAGCGCTTAATAGAAAGCGGTGAGGCAATCTGCTGACGGCG 782	
DB 810 TTTTATTTGACGGTGAACCTGTAGCGAGTTCAATAAGGGTTAAAGTGTCTGTGACGCGTG 869	
QY 783 GTTGTCTCCCGGTTTAAGTCTTAATGACGATGTTGGCAATCCATGCGAGCATTCGCG 842	
DB 870 GTTGTCCGCGGATTTGTGCTTAATCGAGACGGTTTGGCAATTCGATTCGATTCACGTCG 929	
QY 843 ATCCCATCAACCAAGAAGGAATGAACCAAAACGTCGCGCGGAGCAAAATGCGTCCGCTC 902	
DB 930 ATGGTATCGATGTTGAAGGGATGAGTTGTACCGTGTGCCGGGAGCAAAATGCTTACGAGGC 989	
QY 903 TTTTGAATTTCCACAAGAAATTAACAATTTCCGTGACTACGTAATCCCAAGGCCACCGGCC 962	
DB 990 TTTCGAAATTTCAAGGAAGAACTTCAAGTTTCAGAGACTACGTCCTCCGAAAGACAAA--- 1046	
QY 963 CCCGAGCCGCGCAAGAAAGGAGATGAAGGCGTTTATATGATGAATCAGAAAGAGATTTAAGA 1022	
DB 1047 -----ACGAAGTCTCGGGATGAAAGGCTTTGTCTGAGAGATCTGAAAAGAAATTTCTCA 1097	
QY 1023 CTGTGGAGCATGTTTATGTTTGGCAGCTTTCTGTGGATATTTGGGGGTGGCCTTGTGCCGC 1082	
DB 1098 CCGTTGATTTACATCTACGCTGTGACGCGCCTTTGGCGCTACTGGGGTGGTCTGTGCTCG 1157	
QY 1083 AGGTCGTGGCTTGCCTGAGAGCACTGTGATTTACAGCAAGTGTCTTACCAAGGCTGTGAGA 1142	
DB 1158 GAGCTCTACTCTTCCGCCCTCAACATATTTGCCGGCAGAGCTCTCGCGGGGCTTAAGT 1217	

Oy	1143	TGACGATGGAGGAAATTGGGGCGGTGGATTAAGATTGTTCTTCATATGAAGTCGGGGCGGTC	1202
Db	1218	TGACGATGGACAGATCTCGCCGTGGATTAAGATTGTGCATACGGGAATGGATTCGTCTGC	1277
Oy	1203	CGGAGGAAGCGCTGAGGAGATATACGAAGGACCTTCATGCTCATTTGGGAAAAGTTGGGATCG	1262
Db	1278	CGGACATGGCGGAATGAGTTTACGAAGGCTTCACCTCATCTTCCAAAAGCTCGGTATG	1337
Oy	1263	ACGGTGTAAAGATTGAGCTTATCCACCTATTGGAGATGTTGTGTGAAGACTATGGAGGA	1322
Db	1338	ACGGCGTTAAAGTTGAGCTATCCACATATATGGAGATGTGTGGAGAAATATGGCGGA	1397
Oy	1323	GAGTGGATTGGCAAAAGGCAATTACAAAGCAATGACCAATATCAATAATTAACATTTTA	1382
Db	1398	GAGTAGACTTGGGCTAAAGCTTACTTCAAGCGCTTAACTTCCACATGATTAAGCAATTTTG	1457
Oy	1383	AAGGAATGGAGATCATTTGCAAAATATGGAACATTGTATACACTTCAATGTTCCTTGGGACGG	1442
Db	1458	ACGGTAACGGCGTATTCGTAGCATGAGCACTGTATATGATTTTCAATGTTCTTGGAMCCG	1517
Oy	1443	AAGCTATCTCTGTTGGTGGTGGTGGTGGATGAGACTTTTGGTGCAGGAGACCCCTCTGGTGATC	1502
Db	1518	AAGCATCTCTATAGTCGTGTGGGTGATGATGACTTTTGGTGCAGGATCATCAGGCGACA	1577
Oy	1503	CAAAAGGTACGTTTTGGCTCCAAAGATGTACATGTTCTATTTGTGCCAACGACAGCTTGT	1552
Db	1578	TAAACGGCACATATTTGGCTCTCAAGAGATGGCACATGGTCCACATGTGCTCTAACAAAGCTT	1637
Oy	1563	GGATGGGGAACCTTCATCCACCCTGACCTGGGATATGTTCATTCACGCCACCCCTGTGGCG	1622
Db	1638	GGATGGGAATTTTCATCCACCCTGATTTGGGACATGTTTCAATGCACACATCCCTGTGCTG	1697
Oy	1623	CCTTTACATGCTGACCTCTCGAGACCATCTGTGGTGGCCGCAATCTATGTATAGTATCTGTGG	1682
Db	1698	AGTTCCATGGCTGTCTCTGTGGCATCTCGGTGGGCCATTACATCAGCAATTTGTGTGG	1757
Oy	1683	GAAGCATTAACCTTGTATCTTTGAAAAAACTAGTCTCTGTATGATGATCTTGGAA	1742
Db	1758	GCCAGCACGATTTGATCTTGTGAACGACATCGTCTGCTGACGGTTGATTTTGAAGGT	1817
Oy	1743	GTGAGTACTATGCACTCCGCACTCGGATTTGTTTGAACACCTTGTGCTATATGGAG	1802
Db	1818	GTGAGCACTATGCACTCCCAACTCTGTACCGTCTCTTGAACACCTCTCTCATGTATGGCA	1877
Oy	1803	AAACTATGCTTAAAGATTTTGGGAATCTCAACAAGTTCACTGGAATGATGGTGCAATTCACAT	1862
Db	1878	AAACCATGCTCAAGATTTTGGAACTTGAACAAGTACACTGGAATTTATGTGAGCATTCACAT	1937
Oy	1863	GCCAAGGAGAGATGTGTCTGTGTAGACACGCGCAACCAATGCTTTTCCAAATACTCAA	1922
Db	1938	GCCAAGGAGAGATGTGTGTGACAGAAACCAGCAACCAATGTCTTCCAAATCGTTA	1997
Oy	1923	AACGATGACATCTCAAAACTTAACCCAAAGACATAGATATGGACACTGAGAAAACCTTA	1962
Db	1998	ACAGGTTAACCCGACCAACAATCTTAAGGACGTTTAAATGAACAGTGGGAACAACCCAA	2057
Oy	1983	TCTCATATGAAGGCGTTAAAAACCTTTGCGCTTACCTATCTTAAAGCAGCAAAAATTCAC	2042
Db	2058	TCTCGCTTGAAGAAAGCTTGAAGAGTTTCTTGTCTGTCTCAAGTCTAAGAAAGCTTGTGT	2117
Oy	2043	TCTCCAAAGCCTCTCAAGATCTTGACATAGCTCTTGACCCATTTGAAATTCGATGACATCA	2102
Db	2118	TGTCTGGACAAAGATGATCTGAGATCACTTTGGAGCGTTTGAAGTTTGAAGTAAATCA	2177
Oy	2103	CTGTTTCAACGATACCAAACTCATCCAACTCTCTACACTTTGGCCCAATTTGGGCGGG	2163
Db	2178	CTGTCTACCAAGTTGTACTATTAAGGAGTATGTTGGTTAGTTTGTCTCCAAATCGGATTTGG	2237
Oy	2163	TGAACATCTTAAACATAGTGAAGCCATCCAAATCTGTGACATATGACATGACATTAAGCT	2222
Db	2238	TTTAACTCTTAAACATTAAGCGGTGCAATTCGATCTTGGGTATCATATGAGGA-----AT	2291
Oy	2223	CAGTTCGAGATTGGTGTCAAAAGGCTGTGTGATGATGCGAGATATTTTGCATGAAAAAACCA	2283

DB	2232	CGGTGAGATTGCGATTGCTGCTGGAGAGTTTCAGGGTTTATGATCAACAGAAAGCTG	2351
QY	2283	GGGCTTGTCTGATTTGATGGGAGATGTTGGCTTCACATATGATGACGACCAATGTCGG	2342
DB	2352	CGAGCTGCAGAAATGATGATGGTGAAGTTGTTGATGTTGATAG---AAGATCAATGGTGA	2408
QY	2343	TGGTTCAGTGGCATGGCC	2361
DB	2409	TGGTTCAGTGGCTTGGTC	2427
RESULT	5		
LOCUS	E24424	2497 bp	DNA
DEFINITION	Raffinose synthase gene and use thereof.		PAT
ACCESSION	E24424		07-FEB-2001
VERSION	E24424.1	GI:13018164	
KEYWORDS	JP 1999215984-A/2.		
SOURCE	unclassified.		
ORGANISM	unclassified.		
REFERENCE	1 (bases 1 to 2497)		
AUTHORS	Eljiro, W.K.O.O.		
TITLE	Raffinose synthase gene and use thereof		
JOURNAL	Patent: JP 1999215984-A 2 10-AUG-1999;		
COMMENT	SUMITOMO CHEM CO LTD		
	OS Glycine max(soybean)		
	PN JP 1999215984-A/2		
	PD 10-AUG-1999		
	PF 12-DEC-1997 JP 1997342899		
	PR		
	PI EIJIRO WYANABE, KENJI OEDA		
	PC C12N15/09, C07K14/415, C07K16/16, C12N1/21, C12N5/10, C12N9/88, PC		
	C12Q1/527,		
	PC C12Q1/68, G01N33/573//C12N15/09, C12R1:91), (C12N1/21, C12R1:19),		
	PC (C12N5/10, C12R1:91), (C12N9/88, C12R1:91), (C12N9/88, C12R1:19),		
	PC C12N15/00,		
	PC C12N5/00, (C12N15/00, C12R1:91), (C12N5/00, C12R1:91) CC		
	Strandedness: Single;		
	CC Topology: Linear;		
	FH Key		
	FT CDS		
FEATURES	source		
	1..2497		
	/organism="unidentified"		
	/db_xref="taxon:32644"		
BASE COUNT	650 a 521 c 673 g 653 t		
ORIGIN			
Query Match	38.5%	Score 968.6:	DB 6: Length 2497:
Best Local Similarity	65.2%	Pred. No. 1e-269:	
Matches 1536:	Conservative	0: Mismatches 784:	Indels 37: Gaps 6:
QY	39	TTCCTTCTCTCACAAATGGCTCTAGTTTAAAAATGGTGGCTCCACAGTAGTTTCAT	98
DB	45	TCTTGAATTCAGTCCATGGCTCCAAAGCATTAAGCAAAACTGAGAACTAAATTCATTTG	104
QY	99	TGTATGGCTTAATGATGATGCTGTCACCGCTTTGCAATGCACGATCGGATTCACACTGTA	158
DB	105	GTCCTTGTCACAGGATATTTGGCTTGTGCATTAACCTTAAGATCAAAATTTCTCGGCA	164
QY	159	ACGGTCATTGTTCTGTCGTCGATGTTCTGAGAACATTTGTGCTTCTCTCCGTAC-217	
DB	165	ACGGCACCCCTTTTCTCACGGAATTCGCCGAAACATTAATAGTACCCCTTCACCCATCG	224
QY	218	-----ACTTCGATAGACAAAGTCCCGGTTTCGGTGGTGGCTTTGTTGATTCG	266
DB	225	ACGCCAAGATGATAGAACAGACGAGACGACGCTGTAGTGTGCTTGTGGGCTTC	284
QY	267	ACGGCTGGAACTGATACCGACATGTTGTTTCATTTGGAAAGCTGAAGATATTCCGT	326

Db 285 AGCGGACGACGCCAGACCCGACACGCTGCTCCCTGGGGAAAGCTCAGAGCAATAAAT 344
 Oy 327 TTATGAGTATTTTTCAGGTTTAAAGTTTGGTGAAGTACACACTGGGTTGGTGAATGTTG 386
 Db 345 TATGAGCAATATCCGGTTTAAAGTGTGGTGACACTCAGCTGGGTGGAGCAAGGAC 404
 Oy 387 GGGATTTGAATCGGAGCTCAAGATTGTGATCTCTTGAAGAAGTCAAG--ATTGAGTGCAC 443
 Db 405 AGCAACTGGACGACGAGACACAGATGATGCTTTCGACAAAAGACACACAGCTCGAGGCC 464
 Oy 444 CGTATGTTTTCCCTTCTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 503
 Db 465 CCTTGTGTTGATCTTCCGATCTCCGATCTCCGATCTCCGATCTCCGATCTCCGATCTCCGAT 524
 Oy 504 ATGATGATCTTGTGATGTTTGTGTCGAGAGTGGTTCGTAAGATTGTTGATGATCTG 563
 Db 525 TGGATGATTTACGTGAGGAGTGTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 584
 Oy 564 TCCGAGTATGTTGTATCTTCACTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 623
 Db 585 TCGGGAGCTGCTTATAGCTCAGCTGGCCATGACCCGATACAGTGTCTTGAAGAGACAA 644
 Oy 624 TGAAGATCGTGGAGACCATCTTGAACCTTTTGGCTTGTGAGAGAGAGAGAGAGAGAGAGAG 683
 Db 645 CTAAAGTCTGTTAGAGATGATTTGGGAGCTTCAAGCTTCTGAGAGAGAGAGAGAGAGAGAG 704
 Oy 684 GTATCGTGACAAATTCGTTGGTGCAGTGGAGCGCTTTTACCTAACGCTTATCTAC 743
 Db 705 TGATCATAGACAGTTTGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 764
 Oy 744 AGGCGGATATAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 803
 Db 765 CAGGTGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 824
 Oy 804 TTAATCGAGATGTTGGCAATTCATCGACAGATTCGATGATTCATCCATC---CAAGAG 860
 Db 825 TATTCGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 884
 Oy 861 GAATGAACCAACCGCTGCGGAGGAGCAAAATGCCCTGCTTTTCAATTCACAGAGA 920
 Db 885 GTATGAGGACCACTCCGACGAGGAGGAGCAAAATGCCATGATGATGATGATGATGATGATGAT 944
 Oy 921 ATATCAAAATTCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 980
 Db 945 ATATCAAAATTCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 989
 Oy 981 GGAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1040
 Db 990 GTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1049
 Oy 1041 TTTGCAATGCTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1100
 Db 1050 TTTGCAATGCTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1109
 Oy 1101 AGGCACTGTGATTCACCAAGTCTTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1160
 Db 1110 AGGCACTGTGATTCACCAAGTCTTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1169
 Oy 1161 CGGAG 1220
 Db 1170 CGGAG 1229
 Oy 1221 TTTGCAATGCTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1280
 Db 1230 TTTGCAATGCTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1289
 Oy 1281 TTTGCAATGCTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1340
 Db 1290 TTTGCAATGCTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1349
 Oy 1341 CATTTTCAAAAGCAATGACCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 1400
 Db 1350 CATTTTCAAAAGCAATGACCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 1409

Oy 1401 CAGTATGAG 1460
 Db 1410 CGAGCATGAG 1469
 Oy 1461 GTGTTGAT 1520
 Db 1470 GCGTAGGAG 1529
 Oy 1521 TCCAGAGGAT 1580
 Db 1530 TCCAGAGGAT 1589
 Oy 1581 ACCCTGAG 1640
 Db 1590 AGCGGAT 1648
 Oy 1641 GAGGAGGAT 1700
 Db 1649 GGGGAT 1708
 Oy 1701 TTTGAG 1760
 Db 1709 TGTCTCAAGAG 1768
 Oy 1761 GAGCTGAG 1820
 Db 1769 GAGCTGAG 1828
 Oy 1821 GGAATTCACAAATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1880
 Db 1829 GGAATTCACAAATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1888
 Oy 1881 GTCTGAGAGAG 1940
 Db 1889 GTCTGAGAGAG 1948
 Oy 1941 CTAAACCAAAAGGAG 2000
 Db 1949 GAGATCTCTCAAGAG 2008
 Oy 2001 AAACCTTTGCGCTTTTACCTTATCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2060
 Db 2009 ATGTTGTTGCTGATATATTTGTTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2068
 Oy 2061 ATCTTGATATGAT 2120
 Db 2069 AATTGGAAGTTTCACTTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2128
 Oy 2121 AACTATCCAAACCTTCTCTACACTTTGCCCCCAATTTGGGCTGGTGAACATGCTTAAACACTA 2180
 Db 2129 TGTCTGCAAAAAGTTAATTTCAATTTGCTCCCAATTTGGGCTGGTGAACATGCTTAAACACTG 2188
 Oy 2181 GTGAG 2240
 Db 2189 GTGAG 2248
 Oy 2241 AAGGCTGTGAGATGAG 2300
 Db 2249 GGGGATGAG 2308
 Oy 2301 GGGAGAGGATGAG 2360
 Db 2309 GGGAT 2365
 Oy 2361 CAATTGATCTTCATCG 2377
 Db 2366 CTAGTCTTCAAAATG 2382

RESULT 6
 E24423 E24423 2746 bp DNA PAT 07-FEB-2001
 LOCUS

DEFINITION Rafinose synthase gene and use thereof.
ACCESSION E24423
VERSION 2.4423.1 GI:13018163
KEYWORDS JP 199215984-A/1.
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1. (bases 1 to 2746)
AUTHORS Eljiro, M.K.O.O.
TITLE Rafinose synthase gene and use thereof
JOURNAL Patent: JP 199215984-A 1 10-AUG-1999;
SUMITOMO CHEM CO LTD
OS Vicia faba(broad bean)
PN JP 199215984-A/1
PD 10-AUG-1999
PF 12-DEC-1997 JP 1997342899
PR
PI EIJIRO MITSUAKI, KENJI OEDA
PC C12N15/09, C07K14/415, C07K16/16, C12N1/21, C12N5/10, C12N9/88, PC
C12Q1/527,
PC C12Q1/68, G01N33/573, C12N15/09, C12R1/91, C12N1/21, C12R1/19),
PC C12N5/10, C12R1/91, C12N9/88, C12R1/91, C12N9/88, C12R1/19),
PC C12N5/00, C12N15/00, C12R1/91, C12N9/88, C12R1/91, C12N9/88, C12R1/19),
Strandedness: Single;
CC Topology: Linear;
FH key
FT CDS
FEATURES
source Location/Qualifiers
1..2746
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 780 a 552 c 621 g 793 t
ORIGIN

Query Match 38.0%; Score 956.6; DB 6; Length 2746;
Best Local Similarity: 66.6%; Pred. No. 3.1e-266;
Matches 1421; Conservative 0; Mismatches 694; Indels 18; Gaps 3;

QY 248 GGTGTTGTTGATTCGACGGGTGGAACCTGATAGCCGACATGTTGTTGATGGG 307
DB 359 GGTGTTGTTGATTCGACGGGTGGAACCTGATAGCCGACATGTTGTTGATGGG 418
QY 308 AACCTGAAGATATTCGATTTAGATTTTCAAGTTTAAAGTTTGGTGGACTACAC 367
DB 419 AAATTAAGGAATCAAAATTCATGAGCATATTCGGTTCAAAGTTTGGTGGACTAC 478
QY 368 TGGTGGTGGGAATGGTGGGATCTGAATCGAGACCTGAGATTCGATTCCTTGAAG 427
DB 479 TGGTGGGGAACAAATGGACAGAACTACACGAAACAAATGTTAAATCTGGACAAA 538
QY 428 TCAGATTCT--TGGTGCACCGATGTTTCTTCTTCGATCGTTGGAGACGTTCCGA 484
DB 539 AAGCACTCCCTCGAGACACCTATGTTCTCTCTCCATTCCTAGAAACACCTTCGA 598
QY 485 ACCTCGATTACGCTGGGATGATGACTTTTGTGCGATTTGTGCGAGTGGTTCGTCG 544
DB 599 ACCTCACTCCCAACCGGTCCTCAACGATCACATAGGATGTCGCGAAGACGGTTCA 658
QY 545 AAGCTTGTATGATCGTTCGCAAGATATGTTGATCTTATGCTGGTGTGATCCGTT 604
DB 659 CATGTCAACCGGTCACGCTTCAAGCATGTCTTACATCTCACTCAACGACCCATAC 718
QY 605 GCATCTGTTAAAGAGCGATGATGATGATGATGATGATGATGATGATGATGATG 664
DB 719 AATTAATTAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 778
QY 665 GAGAGAAAGATCTCCAGGATATGATGATGATGATGATGATGATGATGATGATG 724
DB 779 GAAGAAAAAAGACACCTCAATATATATAGACAAATTCGGTTGGTTCAGTGGATG 838
QY 725 TACTTAACGGTTTATCAACAGGCGCTAATAGAACGCTGAGGATCTCTGTCGACGG 784

DB 839 TACTTAACGGTTTATCAACAGGCGCTAATAGAACGCTGAGGATCTCTGTCGAC 898
QY 785 TGTCTCCCGGTTTATGCTTAATGACGATGTTGGCAATTCATCGACATTCGGAT 844
DB 899 TGTCTCCCGGTTTATGCTTAATGACGATGTTGGCAATTCATCGACATTCGGAT 958
QY 845 CCCATCCCAAGAGAAAGAAACCAACCGTCCGCGGAGCAAAAGCCCTGCGCTT 904
DB 959 GATGAGATGATTCAGAAATGAACCGAATCTCAGCCGCGGAGCAAAATGCTATGAG 1018
QY 905 TTGAATTTCCAAAGATTAATCAAAATTCCTGATACCTCAACGACCCAGCCGCC 964
DB 1019 GTAAATTAAGCAAGATTCCTAAGTTAGAAATATGAGATCTCTGA-----A 1066
QY 965 CGAGCCGCGCAAGAGGAGATGAAGGCTTTATATGATGATCAACAGAGATTAAG 1024
DB 1067 AATGAGGGAAGAAAGGTTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1126
QY 1025 GTGAGCATGTTTATGTTGGCATGCTTTGTGAGATTAATGGGGTGGCTTCGCGC 1084
DB 1127 GTGAGAGTGTATGTTTGGCATGCGCTTGTGGATTTGGGGGGGGGTTAGGCTTGA 1186
QY 1085 GTGCTGGCTTGGCTGAGGACGTTGATTCAGCAGCTTTCACCGAGGCTGACAGT 1144
DB 1187 GTGATGAGGATGCGCAAGGATGAGGTTGTTTCCGAAGGTCTCAAGGGTTGAAGAT 1246
QY 1145 ACGATGAGGATTTGGCGGTGATAGATTTGTTCTATTAAGTGGGCTGCTCCG 1204
DB 1247 ACGATGAGGATTTGGCGGTGATAGATTTGTTCTATTAAGTGGGCTGCTCCG 1306
QY 1205 GAGAGGCTGAGAGATGATGAGAGGATTCAGCTCATTTGGAAGAAATGGATGAC 1264
DB 1307 GATTTGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1366
QY 1265 GGTGTTAAGATTTGATTCACCTATTTGAGATGTTGTGAGAGATTTGAGAGGAG 1324
DB 1367 GGTGTTAAGATTTGATTCACCTATTTGAGATGTTGTGAGAGATTTGAGAGGAG 1426
QY 1325 GTGATTTGGCAAGGATTTACAAAGCAATGACCAATCAATTAATTAATTTAA 1384
DB 1427 GTGAGCTACCAAGGATTTACAAAGCAATGACCAATCAATTAATTAATTTAA 1486
QY 1385 GGAATGAGTCAATTCAGATGATGAGATTTGATGATGATGATGATGATGATGAT 1444
DB 1487 GGAATGAGTCAATTCAGATGATGAGATTTGATGATGATGATGATGATGATGAT 1546
QY 1445 GCTATCTCTGTTGGTGTGTTGATGATGATGATGATGATGATGATGATGATGAT 1504
DB 1547 GCTATCTCTGTTGGTGTGTTGATGATGATGATGATGATGATGATGATGATGAT 1606
QY 1505 AAGGATGATTTGGTGTGTTGATGATGATGATGATGATGATGATGATGATGAT 1564
DB 1607 AAGGATGATTTGGTGTGTTGATGATGATGATGATGATGATGATGATGATGAT 1666
QY 1565 ATGGGGAATTCATCCACCTGATGATGATGATGATGATGATGATGATGATGAT 1624
DB 1667 ATGGGGAATTCATCCACCTGATGATGATGATGATGATGATGATGATGATGAT 1726
QY 1625 TTGCATGCTGCTGCTGAGGATCTGTTGGTGGGCGGATGATGATGATGATGATGAT 1684
DB 1727 TTGCATGCTGCTGAGGATCTGTTGGTGGGCGGATGATGATGATGATGATGATGAT 1786
QY 1685 AAGCATTAATTTGATTTCTGAAAAAATAGTGTCTCTGATGATGATGATGATGAT 1744
DB 1787 AATGCAATTTTCAAGTTGTCGCAAAATCTGTTTGGCCGATGTTCTATCTGCTGT 1846
QY 1745 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1804
DB 1847 CAACATTAACGACCTCCCTACAAAGATTTGTTTAAAGCCCTTTGATTAAGGCAA 1906
QY 1805 ACATGCTTAAAGATTTGGAATCTCAACAGTTCACCTGAGATGATGATGATGATGAT 1864

Db 1907 ACAATGCTGAAAAATTTGGAAATCTCAACAATAATATACAGGTGTTTGGGTCTTTTCAACTGC 1966
OY 1865 CAAGGAGGAGATGCTGTGCTGAGACACGCCGCAACCAATGCTTTTCAATACATACTAAAA 1924
Db 1967 CAAAGTGTGGTGGTGTCTCTGAGGACGCGCAACAGAGTGTATCTGAATTTTTCACGC 2026
OY 1925 CGAGTACATCCAAAACCTAACCCAAAAGACATGTGAATGGCAGCTGAGAAAACCCATC 1984
Db 2027 GCGGTGACATGTTATGCAAGTCCCGAAGACATTTGAATGGTGCATATGGGAAATCCCAATG 2086
OY 1985 TCGATTGAAGGCTTAAACCTTTGGCTTTTACCTCTATCAAGCAAAAACCTTATCTC 2044
Db 2087 AGCACCACCAAGGTGGATTTTGTGCTGTATATTTTTCAGAGGAGAAATTTGAGCTC 2146
OY 2045 TCCAGCCCTCTCAAGATCTTGACATAGCTCTTGACCCATTCGATTCGATTCATCT 2104
Db 2147 ATGAAGTCTCTGATATGATTTGAAGTTTGGCTTGAGCCATTTAGTTTGAAGCTAATGACA 2206
OY 2105 GTTTCACCACTGACCAAACTCATCCAACTTCTACACTTTGCCCAATTTGGGCTGTG 2164
Db 2207 GTGTCTCCAGTGAAGAGTGTGTTTTCGAAAAGGTTTATACAGTTTGACCGATTTGGTTAGTG 2266
OY 2165 AACATGTTTAACTAGTGGAGCCATTCGATCTGATGACATGACATGACCTAAGCTCA 2224
Db 2267 AACCTGTGAACCTGTGTGTGCTGATTCAGTCTCTGAGAGTTTGAATGCAAGTTTG 2326
OY 2225 GTGAGATGTTGTTCAAAAGGTTGTGTGATGAGATGCGAGTATTTCATCGAAAAACAAG 2284
Db 2327 GTCAAGATTTGGGGTGAAGAGTGTGCGGGAGATGACGTGTTGCTGAGAAACCGGTT 2386
OY 2285 GCTTGTGTTATGATGGGAGAGATGTTGGGTTCAAGTATGATCAGACCAAAATGTTG 2344
Db 2387 TGCCTGCAAAATTTGATGGGTTTAAAGTGAATAATTTCTTATG--AGGACAAATATGGCAAGA 2443
OY 2345 GTTCAAGTGGCCAGCCAAATGATTTCTTCATCG 2377
Db 2444 GTTCAAAATTTGTGGCTTAGTTCTTCAACATTTG 2476

RESULT 7
E24425 1762 bp DNA PAT 07-FEB-2001
LOCUS E24425
DEFINITION Rafinose synthase gene and use thereof.
ACCESSION E24425
VERSION E24425.1 GI:13018165
KEYWORDS JP 1999215984-A/3.
SOURCE unidentifed.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 1762)
AUTHORS Eljiro,W.K.O.O.
TITLE Rafinose synthase gene and use thereof
JOURNAL Patient: JP 1999215984-A 3 10-AUG-1999;
SUMITOMO CHEM CO LTD
COMMENT OS Stachys sieboldii(Chinese artichoke)
PN JP 1999215984-A/3
PD 10-AUG-1999
PF 12-DEC-1997 JP 1997342899
PR
PI FUJURO WATANABE, KENJI OEDA
PC C12N15/09,C07K14/415,C07K16/16,C12N1/21,C12N5/10,C12N9/88, PC
C12Q1/527,
PC C12Q1/68,G01N33/573//C12N15/09,C12R1:91),(C12N1/21,C12R1:19),
PC (C12N5/10,C12R1:91),(C12N9/88,C12R1:91),(C12N9/88,C12R1:19),
PC C12N15/00,
PC C12N5/00,(C12N15/00,C12R1:91),(C12N5/00,C12R1:91) CC
Strandedness: Single;
CC Topology: Linear;
FH key Location/Qualifiers
FT peptide 2..1762.
FEATURES
Source 1..1762
/organism="unidentifed"

BASE COUNT 395 a /db_xref="taxon:32644"
ORIGIN 451 c 499 g 417 t
Query Match 35.2%; Score 886.8; DB 6; Length 1762;
Best Local Similarity 70.0%; Pred. No. 5..3e-246;
Matches 1223; Conservative 0; Mismatches 517; Indels 6; Gaps 2;
OY 379 AAATGTGGGATCTTAAATCGAGAGACTCAGATTTGATCTCTTGAGAGACAGATTCTGG 438
Db 4 AAACGGGTGCGATCTTACGGCGAACTCAAAATGATGCTGTGACACAGTCCGA--CGA 60
OY 439 TCGACCGTATGTTTCTCTTCCGATCGTTGAGGAGCCGCTTCGAACCTCGATTACGCC 498
Db 61 CAGGCCCTTACATCGTGTGCTTCCGCTCATCGAGGGGACATTTGGGCTTCCCTTACGCC 120
OY 499 TGGGATGATGACTTTTGTGATGTTTGTGAGAGTGTGCTGAAATGTTGTTGATGC 558
Db 121 CGGTGTGATGATTTTATGCAATATTGTGTGCAAAAGCGGGTCAACCAAGTCAAGATC 180
OY 559 ATCGTTCGAAGTATGTTGTTATCTTCATGCTGATGATGATCCGTTTGCATCTTTAAAGA 618
Db 181 CTGCTTCCGTGCTTCCGCTTACATGACAGCGCGGTGATGACCTTTTACCTGTGAAAGA 240
OY 619 GCGGATGAGATCGTGAAGACCATCTTGGAACTTTTCGTTGTGAGAGAAAGACTCC 678
Db 241 CGCGTGAAGTTGCGCGCGCCACACCTCGGAGCTTCAAGCTGTGAGGAGAAACTCC 300
OY 679 ACCGATATCGTGAACAATTCGTTGTGTCAGTGGAGCGGTTTAACTTAACGTTCA 738
Db 301 GCGGGGATCGTCAACAATTCGGGTGTGTCACGTGGGATCGTTTACCTCAACGTTCA 360
OY 739 TCCACAGGCGCTTATAGAGGCGTGAAGCATCTGTCAGCGCGGTTTCTCCGGTTT 798
Db 361 GCCCGGATCGTCAACAATTCGGGTGTGTCACGTGGGATCGTTTACCTCAACGTTCA 420
OY 799 AGTCCATATCGACATGTTGTCATTCGATCGACACGATTCGATCCATCCACAAAGA 858
Db 421 GGTGTTGATCGACAGCGGTGTGCACTGATTTGTTCAGCAACAGCGGCTCACACCGA 480
OY 859 AGAATGAACCAACCCGTGCGCGGAGCAAAATGCGCTGCTTTTGAATTTCCAGA 918
Db 481 GGGGATGGGAGAACCTCCCGGAGAGCAAAATGCTTCGAGTTGATCAAGTTTGAAGA 540
OY 919 GAATTACAAATTCGCTGACTACGTCATCCCAAGGCCACCGGCCCCGAGCGGCGAGAA 978
Db 541 GAATTAAGTTACAGGAGGTACGAGAGCCCGATTAACCTGGGCC---GGGCCGAATAC 597
OY 979 GGGATGAAGCGTTTATAGATGAATCAAGAGAGATTAAAGCTGTGAGCATGTTTA 1038
Db 598 GGGATGGGGGCTTTATTCGACATGAGAGACAATTTCAAGAGTGTGACTACGTGA 657
OY 1039 TGTGTCATGCTTTGTGTGATATTGGGGTGGCTTGGCCCGGAGAGTCCGTGGTTGCC 1098
Db 658 CGTGTGATGCGTGTGTGTTTATTTGGGGCGGCTCAGGCCCAATGTTCCGGGCTGCC 717
OY 1099 TGAGCAGCTGTGATTCAGCACTGCTTTCACAGGCGTGCAGATGAGCATGAGATT 1158
Db 718 CGAGGCTAAGCTCATTTAGGCCCAAACTGACTCTGGCTTAAAGACCATGGAAGATT 777
OY 1159 GCGGTGATGATGTTTCTTCATAGTGGGCTGTGCCCGGAGAGAGGCTGAGA 1218
Db 778 GCGTGTGATGATTTGATCAACAATGCGGTGGTCTGGTCCACGAGATTTGTTGAACA 837
OY 1219 GATGTAGAGACTTCATGCTCATTTTGGAAAAAGTTGGGATCGACGTTTAAATTTCA 1278
Db 838 AATGTATGAAGATTACATTCATCTGAATCTGTGGGATTTGATGCTCAAGTTGA 897
OY 1279 CGTATCCACCTATTGGAGATGTTGTGTGAACATATGAGGAGAGGAGTGGATTTGGCAAA 1338
Db 898 CGTATCCATTTTGTGGAATGTTGTGTGAAGACTATGTTGGAGTGGAGTGTGACTTACCAA 957

Db	709	GAGGGAACGCCGTATCGCACATGAGACATGTAATACATTCATGTCCTGGAAAC	768
Oy	1442	GAAGCATCTCTCTGTTGGTGGTGGATGATTTGGGACGAGACCCCTTGATGAT	1501
Db	769	GAAGCATCTCTCTAGTGTGTGCGGTGATGATGATTTGGGACGAGATCATCTGGGAC	828
Oy	1502	CCAAACGGATGCTTTGGCTCAAGAGATGTCACATGATTCATTTGGCCAAACGACCTTG	1561
Db	829	ATTAACGGACGATTTGGCTCAAGAGATGTCACATGATTCATGCTGCTCAACAGCTTG	888
Oy	1562	TGATGGAAGCACTTCAACCCCTGATGGGATGATTCACATCCACCCCTTGCGC	1621
Db	889	TGATGGAAGCACTTCAACCCCTGATGGGATGATTCACATCCACCCCTTGCGC	948
Oy	1622	GCTTCCATGCTCTGAGCATCTGAGCCCATCTGAGCCCATCTATGTTAGTATTCGTG	1681
Db	949	GAGTCTCATGCTCTCTGAGCATCTGAGCCCATCTGAGCCCATCTATGATTCGTG	1008
Oy	1682	GAAGCATTAACCTTGAATCTTCAAAAACTAGTCTGCTGATGATGATCTTCGA	1741
Db	1009	GCGACGACGATTTGATCTCTGAGGAGACGCTGTTTGCCTGACGGTTCATTTGAGC	1068
Oy	1742	ACGATGATCTATGATCTCCGACTCGGATGTTGTTGTTGAACCCCTTGCAATAAGA	1801
Db	1069	TGAGATCTATGATCTCTCCGACTCGGATGTTGTTGTTGAACCCCTTGCAATAAGA	1128
Oy	1802	GAAGCATTAACCTTGAATCTTCAAAAACTAGTCTGCTGATGATGATCTTCGA	1861
Db	1129	AAACCATGCTCAAGATTTGGAACTTGAACAATGATGATGATGATGATGATGATG	1188
Oy	1862	TGCAAGGAGAGAGATGCTGCTGAGACACGCGGACCAACGATCTTTCACATCTCA	1921
Db	1189	TGCAAGGAGAGAGATGCTGCTGAGACACGCGGACCAACGATCTTTCACATCTCA	1248
Oy	1922	AAACGATGATCTCAACCAACTAACCACCAAGCATGATGATGATGATGATGATGATG	1981
Db	1249	AAACGATGATCTCAACCAACTAACCACCAAGCATGATGATGATGATGATGATGATG	1308
Oy	1982	ATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	2041
Db	1309	ATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1368
Oy	2042	CTCTCCAGCCCTCTCAAGATCTTGAATGATGATGATGATGATGATGATGATGATG	2101
Db	1369	CTCTCCAGCCCTCTCAAGATCTTGAATGATGATGATGATGATGATGATGATGATG	1428
Oy	2102	ACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	2161
Db	1429	ACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1488
Oy	2162	GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	2221
Db	1489	GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1542
Oy	2222	TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	2281
Db	1543	TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1602
Oy	2282	AGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	2341
Db	1603	GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1659
Oy	2342	GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	2401
Db	1660	ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1710
Oy	2402	TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	2423
Db	1711	TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1732
RESULT	9		

PSA311087	LOCUS	PSA311087	2727 bp	mRNA	PLN	05-MAY-2001
	DEFINITION	Pisum sativum mRNA for stachyose synthase (stsl gene).				
	ACCESSION	AJ311087				
	VERSION	AJ311087.1	GI:13992584			
	KEYWORDS	stachyose synthase; stsl+ gene.				
	SOURCE	pea.				
	ORGANISM	Pisum sativum				
		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
		Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;				
		Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae;				
		Pisum.				
	REFERENCE	1 (bases 1 to 2727)				
	AUTHORS	Peterbauer, T.				
	TITLE	Molecular cloning of a cDNA encoding for stachyose synthase from				
	JOURNAL	pea seeds				
	AUTHORS	Unpublished				
	TITLE	2 (bases 1 to 2727)				
	JOURNAL	Peterbauer, T.				
		Direct Submission				
		Submitted (02-MAY-2001), Peterbauer T., Institute of Ecology,				
		University of Vienna, Althanstrasse 14, A-1090 Vienna, AUSTRIA				
	FEATURES	Location/Qualifiers				
	source	1..2727				
		/organism="Pisum sativum"				
		/cultivar="Munder von Kelvedon"				
		/db_xref="taxon:3888"				
		/tissue_type="seeds"				
		/dev_stage="developing seeds"				
		/country="Austria"				
	gene	57..2618				
		/gene="stsl1"				
		57..2618				
	CDS	/gene="stsl1"				
		/EC_number="2.4.1.67"				
		/codon_start=1				
		/evidence=experimental				
		/product="stachyose synthase"				
		/protein_id="CAC38094.1"				
		/db_xref="GI:13992585"				
		/translation="MAPPLNSTSLNLIKTESIFDLSEKKFKVKKPPLFDHVDENVSFR				
		SESSICKPSESNAPSLDLOKVLAYSHKGFEGFSGHETPDRIMNSIGSNGRDFLSIF				
		REFTWMSIOWIGSGSDLOMOWILIEVEPKSYVYIPIIEKCFRSLAFGNDP				
		KIIAESGSKVKVESFPNSTAYVHSESNPDLMKKAYSAIRVHLSRFLIEETIPILV				
		DKFQKTPAFITLTPNPICITPHGLDPSGCVPEPRVITIDDCMOSISFPGYDPNEDAK				
		NLVLGEQMSGRILHRPDECFKRTESGLLGPNSPPTPPNNFTDILIKETHEKLR				
		KREBAISKSSDLAEIESKIKRVKEIDDLFGEDEPSGSEKSEMSSEGLAKAFYDRL				
		TFKGLDYYVNHALCGANGVAPETTHDLTKIVPCKLSPLGDTMEDLAVEIISKAS				
		LGIHVHSONEILYDSHMSYLAESGLTGVAVDYISLSEYCDDEYGGVADLAKYIEGLT				
		KSIYKFNFGNMIASMOHNDPEFLGTKOISMGVDDPDPDPNDPNDPMSFMLQGVH				
		MHGSYNSLMKGMOTOPDMDOSDHVCAKPHAGSPATCGGPVYSDNYGSHDPLIK				
		KLVFPDGTIPKCIYFPPLTRDCLPKRPLRDPDTHVLKINPNKIKGCIYIGAPNCGAGMD				
		PLMKRGRFPECKPPIPGVHVHTEVMDKEETSHLGRKEEYVYILNQEISLMTLK				
		SEPIQFTIOPSTFEIYSFVPTKLGIGIFAPIGLIMNPSGTVIDLEYVNGAATIK				
		VKGGGFLAYSSSPKPKFOLNCGEVEFEMVLGDKLGVNVPWIEACGVSDMEIFF"				
	BASE COUNT	809 a	417 c	584 g	917 t	
	ORIGIN					
	Query Match	16.1%;	Score 406.2;	DB 8;	Length 2727;	
	Best Local Similarity	57.7%;	Pred. No. 2.4e-106;			
	Matches	808;	Conservative	0;	Mismatches 578;	Indels 15;
					Gaps 4;	
Oy	975	AGAAAGGAGATGAAAGCGCTTTATAGATCAACTCAAGAGAGATTAAAGCTGTGAGAGATG	1034			
Db	1198	AGTATGGATTGAAAGGCTTTTCACAAAGGATTTGAGGACAAAGTTCAAGGTTTGAGATGATG	1257			
Oy	1035	TTTATGTTTGGCATGCTTTGAGGATGATGAGGATGAGCTTCGCGCCAGAGTGGCTGGCT	1094			
Db	1258	TTTATGTTTGGCATGCACTATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATG	1317			
Oy	1095	TGCGTGAAGCAGCTGTGATTCAGCCAGTGTTCACAGGAGGCTGCAGATGACGATGGAGG	1154			

Db 1318 TTGATACAAAATTGTCCTTGSCAAA---CTCTCACCTGGTCTTGATGGAACAATGAGAG 1374
Qy 1155 ATTTGGCGGTGATAGATTGTTCTTCATAAGGTCCGGCTGGGTCCCGCGGAGAAAGCTTG 1214
Db 1375 ATCTTGCACTGGTGGATGATTTCCAAAGCGTCACCTGGGGAGTGTCAATCCAGTCAAGCTA 1434
Qy 1215 AGGAGATGATGAAAGAGACTTCATGCTCATTTGGAAAAAGTGGAGTGCAGCGGTGTAAAGA 1274
Db 1435 ATGAGCTTTATGATGATCATGATCTTATCTTGTCTGAATCCGGTATCCGGGATCAAG 1494
Qy 1275 TTGACCTTATCCACTATTGAGATGTTGTGTGAAGACTATGAGGAGAGAGTGTGG 1334
Db 1495 TTGATGTCATTCATCTCTTGATATGTGTGTGAATATGATGAGAGTGAAGTTCCTTG 1554
Qy 1335 CAAAGCATATTACAAAGCAATGACCAATCAATTAATAAATTTTAAAGAAATGAGAG 1394
Db 1555 CAAAGTTTACTGATGAAGAGTTGCAAAATCCATGTCTCAAGAAATTTTAAATGAAATGGA 1614
Qy 1395 TCAATGCAAGTATGAAACAATGTAAGAGCTTCATGTTCCCTGGCAGGAAAGTATCTCTC 1454
Db 1615 TGAATTCCTAGCATGACACCTGTAATGACTTTTCTTGTGGAACAACCAAAATTTCTA 1674
Qy 1455 TTGCTGCTGTTGGTGAATGATTTGGTGCAGGACCCCTCTGGTGAATCCAAAGCTACGT 1514
Db 1675 TGGGAAGAGTGGGGATGATTTTGGTCCAAAGATCCAAATGGTGAATCCAAATGGGAAGT 1734
Qy 1515 TTGCTGCCAAGAGATGTCACATGTTTCATTTGTCACACGACAGCTGTGGATGGGAAGT 1574
Db 1735 TTGCTGCCAAGAGATGTCACATGTTTCATTTGTCACATGTTTATGATGGGAGCAAA 1794
Qy 1575 TCAATGCAAGCTGATGAGATGTTGTTGGAAGACCAACCCCTTGCGCGCTTCATGAGCTG 1634
Db 1795 TGAATTCCTAGCATGACACCTGTAATGACTTTTCTTGTGGAACAACCAAAATTTCTA 1854
Qy 1635 CCTCTGAGAGCATCTGTGTGGCCGCAATCTATGTTAGTGTGTGGGAAACATCAAT 1694
Db 1855 GTTCAAGAGATATTTGGTGGGCAATCTATGATGATGATGATGATGATGATGATGATGAT 1914
Qy 1695 TTGATCTTCTGAAAAAGTATGCTCTGATGATGATGATGATGATGATGATGATGATGATG 1754
Db 1915 TTGATCTTCTGAAAAAGTATGCTCTGATGATGATGATGATGATGATGATGATGATGATG 1974
Qy 1755 CACTCCGAGTCCGAGATGTTGTTGGAAGACCCCTTGCAATTAAGGAACATGATGCTTA 1814
Db 1975 CACTTCACACCTAGAGATGTTCTTTCAAAAACCCCTTATTCGACCATACAGTCTCTCA 2034
Qy 1815 AGATTTGGAATCTCAACAAGTCTCATGAGTGTGTCATTCACACTGCCAAGAGAGAG 1874
Db 2035 AATTTTGGAACTTCAACAAGTATGAGAGCGGTGATGTTGCTTCAACTGTCCAAAGGGCAG 2094
Qy 1875 GATGCTGCTGTGACACAGCCGCAACCAATGCTTTTTCACAATCTCAAAAAGAGTACAT 1934
Db 2095 GGTGGGATCCAAATTAAGCAAGATTTAGGGGCTTCCCTGATGCTACAAACCAATACCCG 2154
Qy 1935 CCAAAACTAACCCAAAAGACATAGATGAGCAAGTGAAGAAAACCCATATCTCATTTGGGTA 1994
Db 2155 GTACTGTTCAATGTAAGCAAGTGAATGGAATCAAAAAGAAACCAATCTCATTTGGGTA 2214
Qy 1995 GCGTTAAACCTTTGGCGCTTACCTCATACAGCCAAAAAAACT---TATCTCTCTCAAGC 2051
Db 2215 AAGCAGAAAGAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2274
Qy 2052 CCTCTCAAGATCTTGAACATGCTCTGAGCAATTCGAATTCGAGCTGATGATGATGATGAT 2111
Db 2275 AATCCGAACCAATTAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2334
Qy 2112 CAGTGAACCAATGATCAAAATCTCTACATTTGGCCCAATTTGGCGCTGGAAGTGC 2171
Db 2335 CCGTTTCAAAA---GTTATGTTGGCGGCAATCAAAATTTGGCACCAAATTTGGGTGACAAACATGT 2391
Qy 2172 TTAACTACTGAGAGCCATCAATCTGTGACTATGAGATGAGTACCTTAAGCTAGTGCAGA 2231
Db 2392 TCAACAGTGTGGAGACGTTATGATTTGGAAATATGTTG-----GAAATGGTGTCTAAGA 2445

Qy 2232 TTGCTGCAAAAGGCTGTGGTGGAGATGCGATATTGATCGAAAAAACCAAGGCTTGTG 2291
Db 2446 TTAGGTTAAAGGTGGTGGAGCTTTTCTGTTATTTCTAGTACATCACCAAGAGATTTTC 2505
Qy 2292 GTATTATGAGGAGAGATGTTGGCTCAAGTATGATGAGACCAAAATGTTGGTGTTCAG 2351
Db 2506 AGTTGAATGCTTGTGATAGTGTGATGAGTGGCTAGTGTGATGGAATAATGTGTGTAATG 2565
Qy 2352 TGCCATGGCCCAATGATCTT 2372
Db 2566 TTCTTGATTTGAAGAGCTT 2586
RESULT 10
VANI9024
LOCUS VANI9024 3073 bp mRNA PLN 25-JAN-2000
DEFINITION Vigna angularis mRNA for galactinol:raffinose
6-galactosyltransferase.
ACCESSION Y19024.1 GI:6634700
VERSION Y19024.1 GI:6634700
KEYWORDS galactinol:raffinose galactosyltransferase.
SOURCE adzuki bean.
ORGANISM Vigna angularis
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Vigna.
REFERENCE 1 (bases 1 to 3073)
Peterbauer, T., Mucha, J., Mayer, U., Popp, M., Glossl, J. and
Richer, A.
Stachyose synthesis in seeds of adzuki bean (*Vigna angularis*):
molecular cloning and functional expression of stachyose synthase
Plant J. 20 (5), 509-518 (1999)
20117502
MEDLINE 2 (bases 1 to 3073)
AUTHORS Peterbauer, T.
TITLE Direct Submission
JOURNAL Submitted (12-FEB-1999) T. Peterbauer, Institution University,
University of Vienna, Institute of Plant Physiology, Althanstrasse
14, A-1091 Wien, AUSTRIA
FEATURES
source location/Qualifiers
1..3073
/organism="Vigna angularis"
/db_xref="taxon:3914"
343..2916
/EC_number="2.4.1.67"
/product="galactinol-raffinose galactosyltransferase"
/protein_id="CA64363.1"
/db_xref="GI:6634701"
/translation="MAPPNDPVNATLGLPESEKVPDLSDGKLTGVKGVLLSHVENV
TFSFSSICVPRAPPSILQRYVAASHKGFGLGSHVSPDRILNSGSRGRLFSIF
REKTMSTOWNGSGDLMETQWILIEVETESYVILIPLEKSRGALHPGSDIH
KTCAGSGTOVRASSFGALAYVAHYAETPNLMEAEASALVRHDSRLLEKVPRI
DKRGCTMDPAFLYLVNPGVGMGLKDFSGGVAAPRVVYIDGQOSVNPDPDEDPDAK
NVLVGEBOUATRLRHEBDDKRYKGLGLGNAPSPERTIKELISGIEAHEHGK
QAAAISAGSDLAIEILMTVKNREEDIDLEFGKRSNDSGCCAAACGGKMDFTT
DLRTEFKGLDDVYVMAHLCGGMGVPRGTHLDSKILPKLSGLVTKMDLAVDKIV
EGSIGLVHPHOANDLDSMHSYLAQGVGVKIDVLSLEYCEYEGVGEVLAAYD
GLNSIIRKNPNSGDIASMOOCNDEFELTKOIPRGVGDDEFWOPNDPMPGVFPI
GVHMTCSYNSLSMGOIOPDMDRQSDHCAKFNHAGSRALICGGPYVVSNSHSD
LTKLVPRPGTVPKCIFFPLPTDCLFRNPLFDQKTVLIAMNKKIGVYIAGPNCGA
GMDPKRKKRKEPECKYKASIVHVEVEDOKREAHMGKAEVYVYVQAEVLLHM
TPVSEPLQLIQIPSTPELVNFEVETLGSNIFAPRIGLNMNSGCTIOBLEYIKD
VKVYVGGGRFLAYSTQSPKKFOLNSDAFQWLPDGLKLTLMVAITEENDGVSDLAIF
F"
BASE COUNT polyA_signal 3021..3026
COUNT 814 a 595 c 737 g 927 t
ORIGIN
Query Match 15.6%; Score 392; DB 8; Length 3073;

Best Local Similarity 55.8%; Pred. No. 3 2e-102; Matches 812; Conservative 0; Mismatches 630; Indels 12; Gaps 3;

980 GGGATGAAGCGCTTTATAGATGAACTCAAGAGATTTAAAGCTGTGAGCATGTTAT 1039
 1498 GGGATGAAGACTTCAACAGAGATTTGAGAGATGAATCAAGATTTGGATGATGTTAT 1557
 1040 GTTGGCATGCTTTGTGTGATATTGGGGGTGGCCCTCGCCGAGAGTGGCTGGCT 1099
 1558 GTGTGGCATGCTGTGTGGCGGATGGGGGTGTGAGGCA--GGAACTACACATCTT 1614
 1100 GAGGCAAGTGTGATGAGCCAGTGTTCACAGGGGCTGAGATGAGATGAGATTTG 1159
 1615 GATTCGAAATTAATACATGCAAACTCTCTGCTGTTTGGGACCATGAAAGATCTT 1674
 1160 GCGGTGATTAAGATTTGTTCTTCAATAGGTCGGCTGGTCCCGGAGAGAGCTGAG 1219
 1675 GCATGTGATTAATAGGAGAGATTTCCATAGGGCTTTTCATCTCATCAAGTAATGAT 1734
 1220 ATGTACGAGGACTTCATGCTCATTTGGAAAAAGTTGGATTCACAGCTGTTAAGATTGAC 1279
 1735 CTCTGATTCACATGACCTTATCTTGTGCCCAACCGGTGTACTGAGAGTCAAAATTTGAC 1794
 1280 GTTATCCACTTATTTGAGATGTTGTGAGAGCTATGAGAGGAGAGATTTGGCAAG 1339
 1795 GTATTCACAGCTTGTGAATAGTATGAGAGAAATGAGAGAGAGATTTGCCAAG 1854
 1340 GCATATTCAAAAGCAATGACCAATCAATTAATTAACATTTTAAAGAAATGAGATCTT 1399
 1855 GCTTATTACATGGGTGAGCAAACTCATATCAAGAAATTTCAATGAGAGTGGATCATC 1914
 1400 GCATATGAGCAATTTGTAACGACTTCATGCTTCTGGCACGGAGCTATCTCTTTGGT 1459
 1915 GCTATGATGCAAGATGCAAGATTTCTCTCTTGTGAACCAAAATTTCCCTTTGGA 1974
 1460 CGTGTGATGATGACTTTGTGTGACGAGACCCCTGCTGATCCAAACGGATTTGG 1519
 1975 AGAGTGTGGATGACTTTGTGTGCAAGATCCGATGGAGGCCAAAGGGAGTCTTCGG 2034
 1520 CTCGAAAGATGTCATGATGTTGTCACAGCAAGCTTTGGATGGGGAATCTTATC 1579
 2035 TTACAGGGGGTGCATGATGATCTACTGTTCTACAGACGTTGTGATGGGACATATATT 2094
 1580 CACCTGACTGGGATGATGTTCAATCCACCACCCTTGTGCCCTTCATGCTGCTCT 1639
 2095 CAACCCGATTTGGGACATGTTCAATCGCATGATGATGATGATGATGATGATGAT 2154
 1640 CGAGCCATCTCTGTGGCCGATGATGATGATGATGATGATGATGATGATGATGAT 1699
 2155 AGGCTATTTGT 2214
 1700 CTTCGAAAAAAGTACTGCTTCTGATGATGATGATGATGATGATGATGATGATGAT 1759
 2215 CTCTATTAAAGATGCTTCTCTCTGATGATGATGATGATGATGATGATGATGATGAT 2274
 1760 CCGACTGCGATGCTTTGTTGAAGACCTTTGATGATGATGATGATGATGATGATGAT 1819
 2275 CCAACAGAGATGCTCTTCTGAGGACCTCTCTTGTGACCAAAAAACCGTTCTCAAAAT 2334
 1820 TGGATGTCAAAGTCTGATGAGATGATGATGATGATGATGATGATGATGATGATGAT 1879
 2335 TGGAACTTCATTAAGATGATGAGATGATGATGATGATGATGATGATGATGATGATGAT 2394
 1880 TGTGATGAGACAGCCGCAACATGCTTTTCAATATCAATATCAATATCAATATCAAT 1939
 2395 GACCAAAAGGGGAAAGATTTAAAGGTTTCCAGAGTGTTCAAAGCATATCATGACT 2454
 1940 ACTAACCAAAAGCATAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 1999
 2455 GTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2514
 2000 AAAACCTTGTGCTTCTATCAAGCCAAAAACT---TATCTCTCCAAACCTCT 2056

Db 2515 GAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2574
 Oy 2057 CAAGATCTTACATAGCTGTTGACCAATTCGAAATTCGAGCTCAATCTGTTCCAGATG 2116
 Db 2575 GAACCACTGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2634
 Oy 2117 ACCAACTCATCCAACTTCTACACTTTGCCCCAATTTGGGCTGGTGAACATGCTTAC 2176
 Db 2635 GAAAGCTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2694
 Oy 2177 ACTAGTGAAGCCATTCATGCTGAGATGATGATGATGATGATGATGATGATGATGAT 2236
 Db 2695 AGTGAAGGACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2751
 Oy 2237 GTCAAAGGCTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2296
 Db 2752 --AAGGCTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2356
 Oy 2297 GATGGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2416
 Db 2809 AATGCTTCTATGCTGCTTTCATGATGATGATGATGATGATGATGATGATGATGAT 2468
 Oy 2357 TGGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2516
 Db 2869 TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2574
 Oy 2417 TTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2634
 Db 2929 GTTCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2694

RESULT 11
 AP003282/c 135295 bp DNA PLN 24-JUL-2001
 LOCUS Oryza sativa genomic DNA, chromosome 1, PAC clone: P0583608,
 DEFINITION complete sequence.
 ACCESSION AP003282
 VERSION AP003282.2 GI:15004914
 KEYWORDS HTG.
 ORGANISM Oryza sativa (cultivar: Nipponbare) DNA, clone: P0583608.
 SOURCE Oryza sativa
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.
 REFERENCE 1 (bases 1 to 135295)
 AUTHORS Sasaki, T., Matsumoto, T. and Yamamoto, K.
 TITLE Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
 clone: P0583608
 JOURNAL Published only in Database (2001) In press
 REFERENCE 2 (bases 1 to 135295)
 AUTHORS Sasaki, T., Matsumoto, T. and Yamamoto, K.
 TITLE Direct Submission
 JOURNAL Submitted (19-FEB-2001) Takuji Sasaki, National Institute of
 Agricultural Resources, Rice Genome Research Program, Kannondai
 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
 (E-mail: tsasaki@nias.affrc.go.jp, URL: http://rgrp.dna.affrc.go.jp/
 tel: 81-298-38-7441, fax: 81-298-38-7468)
 COMMENT On Jul 24, 2001 this sequence version replaced gi:13072312.
 The orientation of the sequence is from T7 to Sp6 of the PAC clone.
 FEATURES
 source
 1. 135295
 /organism="Oryza sativa"
 /cultivar="Nipponbare"
 /db_xref="taxon:4530"
 /chromosome="1"
 /clone="P0583608"
 BASE COUNT 38342 a 30200 c 29453 g 37300 t
 ORIGIN
 Query Match 14.5%; Score 364; DB 8; Length 135295;
 Best Local Similarity 60.2%; Pred. No. 1.1e-93;
 Matches 729; Conservative 0; Mismatches 415; Indels 66; Gaps 5;

Oy	115	CATGCGCACCGTTTGCATTCGACGAGATTCGATTTCACTGGAACGGCATTCGTTCT	174
Db	9272	CMACCCGCCGGGTACGCTCAAGGCAAGACCTGCCGTGAGAGGGCACCCCTCTCT	92133
Oy	175	GTCGATGTTCTCGAGAACATTTGTCTTCTCTCTTCTCCGTAC-----ACTTTCATAGA	228
Db	9212	CCCTGACGTGCGGGCAAACTCCGGCCCTCAACCCCGGGGTCCACCCCTGATCCCACTCCGA	91535
Oy	229	CAAGTCCCCGGTTTCGGTTGGTTCGTTTGTGGATTGACGGGTGGGAACCTGATAGCCG	288
Db	9152	CGTCCCGCCGGCCGGCCCGCCGACGCTTCTCGGCTTTCGACGGCGGGGCCCAAGACCG	90933
Oy	289	ACATGTTGTTTCGATTTGGGAAAGCTGAAAGATATTCGTTTATAGATATTTTCAGSTTAA	348
Db	9092	CCACGTGGTCCCATCGGTAAAGTCCCGGCACACAGAGTTCAATAGACATCTTCGTTTCA	90333
Oy	349	GCTTGGTGGACTACACACTGGGTTGGTGCAGAAATGGTGGGATCTTGAATCGGAGACTCA	408
Db	9032	GGTGTGGTGGACGACCCACACTGGGTGGGTACAAACGCGCAGGACGTGGAGAACGAAACCA	897373
Oy	409	GATTGTATCTTTGAGAAAGTCAGATTTC-----TGGTCAGCCGTATGT	450
Db	8972	GATATATATCTTACACAGTCGTGGGAGAAAGTCTTGCCGACGGCCCTCGGCCCTTAGGT	89133
Oy	451	TTTTCTTTCCTTCGATCTGTTAGGGACCGTTTCCGAACTCCGATTCAGCCCTG--GATAGA	507
Db	8912	CCCTCTCTCTCCGATTCGTGAGGGGGCCCTTCCGGGCGTCCGTGAGTCCGGGAAAGCCGA	88535
Oy	508	TGACTTTGTGCAGTGTGTTGTCTGACAGAGTGTTCGTGCAAAATTGTTGATGATCTGCTCG	567
Db	8852	GGACTAGCTGTGACATGTGCTCTCGAGAGCGGGTGTCTCACAGTCAGAGGGCTCCGTGTTCCG	87933
Oy	568	AAGTATGTGATATTTATCATCTGCTGTGATGATCCGTTTGCACCTGTTAAAGAGCGCATGA	627
Db	8792	GAGCGCTGTGACTGCAACCGGGGAGACACCCTTTGACTGTGTCAAGAGCCGATGAG	87333
Oy	628	GATCGTAGAGACCCTATCTTGGAACTTTTCCGTTTGGAGGAGAAAGACTCCACAGTAT	687
Db	8732	GGTGTCCGTGCTCCACCTCGGCACGTTTCCGCTCATGAGAGAAACCCCGCCCGCAT	867373
Oy	688	CGTGACAAATTCGTTGGTGGAGCATGCGGAGCGGTTTACCTAAACGTTCAATCCACAGG	747
Db	8672	CGTGACAAATTCGTTGGTGGAGCATGCGGAGCGGTTTACCTAAAGTCCACCCGAGGG	86133
Oy	748	CGTATATGAAAGGCTGAGGCAATCTGTCGACGGCGGTTGCTCCCGGTTTATGCTCTAT	807
Db	8612	GGTGTGGAGGGCGTGCGGGGGCTCGGCACGGCGGGGTGCCCGGGCTGTGTGAT	85535
Oy	808	CGACGATGTGTGGCAATCCATCGGACACGATTTCCGA-----TCCATATCCAAAGAGG	861
Db	8552	CGACGAGGGCTGGAGTCGATCTGCCACGACGACGACCTGGGAGGGGCGCGAGGG	84935
Oy	862	AATGAACCAAAACGCTCCCGCGGAGCAAAATGACCTGCTTTTGAATTTTCAAGAGAA	921
Db	8492	GATGAACCGCACGTCGGCGCGGACGACGATGCCGTGACAGGTGATTAATTTCCAGAGAA	84333
Oy	922	TTACAAATTCGCTGACTACGTCAATCCCAAGGCCACCGGCCCGGACCGCCAGAGGG	981
Db	8432	CTACAAATTCGAGGAGTA-----CAAGGGCGCG	8406
Oy	982	GATGAAGCCGTTTATACATGAACTCAAGAGAGATTTAAGACTGTGGAGCATGTTATGT	1041
Db	8405	CATGGCGCGCTTCTGTAGAGGATGAAGCGCGGCTTCCGACGGTGTAGACAGTGTAGT	8344
Oy	1042	TTTGGATCTTTGTGTGATATTTGGGTGGCTTTCGCCCGCAGTGCCTGCTTGCTGA	1101
Db	8345	GTTGGACCCCTGTGCGGGTACTGGGGGCGCTCCGCCCGCGGCGCCGGGCTTGGCGCC	8288
Oy	1102	GGCACGTGTATTCAGCAGTGCCTTTCACAGGGGCTGCAGATGACATGGAGATTTTGGC	1161
Db	8285	GCCGAGGTCTTCCCGCAGAGCTCTCGCGGGCTTGCAACGCCACCATGTAGAGACTCGC	8222

[illegible]

Db 55108 CATTCGACGATTTATGTCTTCCTTGGACGAGCATCTCTTGTCGTGTCGTAAG 55167
Oy 1464 ----- 1463
Db 55168 TCTTTTACTTTCTGTTTAAATGAACCTTAAGACTATTTTCAAAATTTTCATATTAAT 55227
Oy 1464 -----TTGGTGATGACTTT 1477
Db 55228 TATTTTACTTAACTAATGATTCATTTGATTTGTTGGAATTTAGGTGATGACTTT 55287
Oy 1478 TGTGTCACGAGACCCCTGTCGTGATCCAAACGGTACGTTTGGCTCCAGATGTCACATG 1537
Db 55288 TGTGTCACGAGATCCATGAGGAGATCAAAACGGTACGTTTGGCTCCAGATGTCACATG 55347
Oy 1538 GTTCATTGTGCCAACGACAGCTGTGATGGGAACTTCATCCACCTGACTGGGATATG 1597
Db 55348 GTCCACTGTGATACMACAGCTCTGATGGGAAATTTTCATTCACCTGATGGGACATG 55407
Oy 1598 TTCCAAATCCACCCACCCCTTGTGCGCCTTCCATGCTGCTCCGAGCCATCTGTGAGC 1657
Db 55408 TTTTCAGTCCACACATCTTGTGCTGATGATTCATGCTGCTTCTCGGGCCATTTCCGGTGG 55467
Oy 1658 CCGATCTATGTTAGTATGATCTGTGGGAAAGCATATCTTGAATCTTGTGAAAAACTAGTG 1717
Db 55468 CCGATTTTACATCAGTATGATGTGGGCAAGCATGATTTTGAATCTTGTGAAAGCGCTTGT 55527
Oy 1718 CTTCCTGATGATGATCTCTGCAAGTATGATGATGATGATGATGATGATGATGATG 1777
Db 55528 TTGCCCAACGGTCTGATTTGATGATGATGATGATGATGATGATGATGATGATGATG 55587
Oy 1778 TTTTGAAGACCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1837
Db 55588 TTTTGAAGATCTCTTCTTCAATGATGATGATGATGATGATGATGATGATGATG 55647
Oy 1838 A 1838
Db 55648 A 55648

RESULT 15
BOPTIN 1669 bp mRNA PLN 04-APR-1995
LOCUS B.oleracea putative imbibition protein mRNA.
DEFINITION X79330
ACCESSION X79330.1 GI:488786
VERSION 1
KEYWORDS Imbibition protein.
SOURCE Brassica oleracea.
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 2669)
AUTHORS Fujikura, Y. and Karssen, C. K.
TITLE Cauliflower cDNA encoding a putative imbibition protein
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2669)
AUTHORS Fujikura, Y.
TITLE Direct Submersion
JOURNAL Submitted (17-MAY-1994) Y. Fujikura, Dept of Plant Physiology,
Agriculture University, Arboretumlaan 4, 6703 BD Wageningen,
NETHERLANDS

FEATURES
source Location/Qualifiers
1..2669
/organism="Brassica oleracea"
/cultivar="Alpha Baloma"
/db_xref="taxon:3712"
/dev_stage="early germination"
/tissue_type="root tips"
/cell_line="pBlueScript SK9(-) in XL1-Blue"
/clone="Op6-1-6"
175..2472
/codon_start=1
/product="putative imbibition protein"

/protein_id="CAA55893.1"
/db_xref="GI:488787"
/db_xref="SpreMBL:043408"
/translation="MTTSNTSVNDLVVQKTIILKIPDNITLTTPVAGSGNSGAF
IGATFKOSKSLHVPFIVLEBGRFAVLQGNKNEIKLESGDAVAGSOTHLVYH
DEVNGDAPVYVYFELPDLBGRFAVLQGNKNEIKLESGDAVAGSOTHLVYH
AGTNPFEVITQSVAAEROMOTFHHREKKLIPSEVDNMGCTMDAFITDVAEVEDG
LRSLSEGCTPPREFLIDGMOQIENKEDSCIVQEGNOFATRLVGIKENAKPOKNDP
KDTQVSLKSVYDNAAKORHNYKQYNAHMLALGAGYKPRASGHEHDSALAYVQSP
GVLSNPDIYVDSLVHGLVNPKNKYFNEILHSLVASCIDGVAVDQNTLETLS
AGLGRVSLRSHYQALEASTARNKDNQNCISCHNTDGLYSAKQATVYASDYP
RDNPSHVIHLASVAVNTLFLGEPQPMDEHSHIPAEYHAAAVAGCGAIVSDP
GNHNEFLRLKVLVDGSLRARLPGRPRDCLFPADPADGSLIKLIMNMKFTIYGV
FNCQAGCKDQTKNRHIDHSPGTLVLRVEDDLISEVAGDMDGSDIYAYKSE
LVRLPKGASIPDLTKVLEIELEFHLISPLKEITASIFRAPICLIDPNSGATQSMETN
VTDEKPELSSSVSENRSPALISLVGRGCRGAIVSSQRLCAVDGETETERNYDA
EVGLVLTINLPTREMRWRVEILY"

BASE COUNT 683 a 577 c 655 g 754 t
ORIGIN

Query Match 10.0%; Score 252.2; DB 8; Length 2669;
Best Local Similarity 51.1%; Pred. No. 1.3e-61;
Matches 834; Conservative 0; Mismatches 718; Indels 81; Gaps 7;

Oy 282 ATAGCCGACATGTTGTTGATGAGGAGCTGAGGATATTTGATGATTTTCA 341
Db 332 AGAGTCTCACGCTGTTCCCATCGGTGTGAGAGGCTTGTGATGTCTGTTTC 391
Oy 342 GGTTTAAGGTTTGTGAGCTACACAGCTGGTGTGCAAAATGTTGGGATCTTGAATCGG 401
Db 392 GGTTTAAGCTATGATGATGATGATGATGATGATGATGATGATGATGATG 451
Oy 402 AGACTCAGATTTGATCTTGTGAGAGTCAAGATTTGTCGACCGATATGTTT----- 452
Db 452 AGACACAGTTCAATGTTACTGAGAGTAAAGATGAGATTAACGAGATGATGCTCTACTG 511
Oy 453 -----TCCCTTCTCGATGTTGAGAGGACCGTCCGAACTCGATCGACCTGGGG 503
Db 512 TTTTATACCGTCTCTCCCTCCCTACTTGAAGCCAGTTCAGACCTGTTTGCAGAGAAATG 571
Oy 504 ATGATGACTTTGTCATGTTGTGTGAGAGTGTTCGTCGAAGTTGTGATGATCGT 563
Db 572 AGAAGACAGAGATGAGATTTGTCTTGAAGTGTGATGATGATGATGATGATG 631
Oy 564 TCCGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 623
Db 632 GAACACACCTGTCATGTCATGCTGGAGCTAATCCCTTGAAGTTATCACGCAATCGG 691
Oy 624 TGAAGATGTTGAGACCCATCTGGAACCTTTCGCTTGTGAGAGGAGAAAGTCCACAG 683
Db 692 TAAAGCTGTCAGAGAGCAAAATGACAGACCTTTCATCACCGTGAAGAAAGAGCTGCTT 751
Oy 684 GTATGTCGACAAATTCGTTGGTGTGACAGTGGAGCGCTTTTACCTAACGTTTCAATCC 743
Db 752 CTTTGTGAGACTGTTTGGATGTGTACATGGAGCGCTTTCACACAGAGCTGATCGCG 811
Oy 744 AGGCGTAAATGAGAGCGCTGAGGACATCTCGTGCAGCGCGCTTCTCCCGGTTTATGTC 803
Db 812 AGGGGCTGATGAGGCGCTTGAAGCTTTCAGAAAGAGAGTCTCCACCAAGGTTTCTGA 871
Oy 804 TAAATGAGAGGTTGGGAAATCCATCGGACAGATTCGATCCATCACCAAGAAAGAA 863
Db 872 TCATGATGATGAGCTTGGCAAAATGAAATGAAAGAG-----AAGATTCA 919
Oy 864 TGAACCAAAACCTCCCGCGAGCAAAATGCCCTGCTTTTGAATTTCAAGAGAAAT 923
Db 920 ACTGCTGTTGTCAGAGAGAGACAGTTTGTCTAGAGCTTTTGTGTAATAAGAGAAAG 979
Oy 924 ACAATTCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 983
Db 980 CAAATTTTCAAAAGAAATG---ATCCTAAGGACACGCAAGTGTCAAGGCTTGAAGTGT--- 1033

QY	984	TGAAGGGTTTATATGAGACCTCAAGAGAGAGCTTTAAAGCTGTGGAGCATGTTATATGTT	1043
Db	1034	-----TAGTGAACATAGCAAGACGGCCATATATGTGAAGCAAGTATGCGT	1081
QY	1044	GGCATGCTTTGTGTGGATATTTGGGGGTGGCCCTTCGCCCGAGGTGGCTGGCTGGCC----	1098
Db	1082	GGCAACGCTTATAGCTGGGCTACTGGGGGGAGATGAAACCCGAGACTTCGGTATGGAACACT	1141
QY	1099	-TGAAGCAGCTGTGATTCACGCCATGCTTTACACAGGGCTGCAGATGACATGAGAGATT	1157
Db	1142	ACGACAGATGCAATTAGCATATCCACAGTCCAGTCCCCAGAGTCTTTAAGGAAACCAACCATATA	1201
QY	1158	TGGGGGTGAATAGATTTGTCTTCATTAAGGTGGGTGGGTGCTCCCGCGGAGAAAGCTGAGG	1217
Db	1202	TAGTAATGAGACAGCTCGCTGCTTCATGAGACTGGCTGCTTTCACCCCAAGAAAGTTTCA	1261
QY	1218	AGATGTACGAAGGACTCATGCTCATTTAGGAAAAATTTGGATACAGAGTGTAAATTTG	1277
Db	1262	ACTTCTACAAAGAGCTGCACCTCTACCTGGCTTCGTTGTGTATTAACGAGATCAAAAGTTG	1321
QY	1278	ACGTTATCCACCTAATTGGAGATTTGTGTGAAGACTATGAGAGGAGAGATTTGGCAA	1337
Db	1322	ATGTTTCAAGACATCATGAAACTCTTGHTGHTGTTGTGGTGGAGAGTCTCTCACTC	1381
QY	1338	AGGCAATTTACAAAGCAATGACCAAAATCAATTAATTAACATTTTAAAGAAATGAGTCA	1397
Db	1382	GCACCTTACCAACAGGCGTTTGAAGCTTCCATGTCTAGAAACTTTAAATATATGGGTGCA	1441
QY	1398	TTTGACAGTATGGAACATTTGAACGACTTCATGCTTCCTGGACGGAGACCTAATCTCTTG	1457
Db	1442	TTTCTTGCAATGTGCACACACGGAATGGTTATACA--GCCGGAACAGACTGCTATTG	1498
QY	1458	GTGCTGTGTGTGAGACTTTTGGTGCACGGACCCCTGTGGTGTATCCAAACGGTACGTTT	1517
Db	1499	TTTAGCGGTGTGATGATTACTATCCGAGAGATCCGTCTGCACACTA-----	1546
QY	1518	GGCTTCACAGATGTGCACATGCTTATTGTGCCACAGACAGCTTTGTGATGGGGAACTTCA	1577
Db	1547	-----TACATTAATAGCATGCGTTGTCTTACAAATCTTTTCCTTGGAGAGTTCA	1594
QY	1578	TCCACCTGCAGCTGGGATATTTTCCAAATCCACACCACCTGTGTCGCCCTTCATGCTGCT	1637
Db	1595	TGCACCTGACTGGGACATTTTCCATAGTTTACACCCCAACTGACAGATACCATGCTGCGG	1654
QY	1638	CTGAGCCATCTGTGTGGCCGCGATCTAATTAGTATTTGTGGGAAAGCAATCTTG	1697
Db	1695	CGGTGTGAGTGGGTGTGTCATCTATGTCAATGATGAACGACCAAGCAACCAACTTG	1714
QY	1698	ATCTTTGAAAAAAATAGTCTCTCCGATGAGANTGCATCTCTTGAAGTATGATATATGAC	1757
Db	1715	ATCTATTGAGGAAGCTGGTCTTCTTCGATGGTTCAGTTCCTGTGCTCGGCTCCGGGTA	1774
QY	1758	TCCCGATCCCGATGTTGTTTGTGAAGACCCCTTTCGATATGAGAAACTATGCTTAGA	1817
Db	1775	GACCTACAGCTGACTGCTATTGTCGATCCAGCTAAGAGATGGAACCAAGCTTGTAAAGA	1834
QY	1818	TTTGAATCTCAACAAGTCTACTGAGATTTGGTGCATCTCAACTGCCMAAGAGAGAGAT	1877
Db	1835	TCTGGAACATGAATAAGTTACTGATATAGTTGGTATATCAACTGTCAAGGTGCTGTT	1894
QY	1878	GGTGTCTGAGAC	1890
Db	1895	GGTGCAGAGATAC	1907

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 1, 2001, 00:48:33 ; Search time 227.5 Seconds
(without alignments)
9485.220 Million cell updates/sec

Title: US-08-846-234-4
Perfect score: 2517
Sequence: 1 AAAAACAACCTCTTTTA.....TTGACAGTAATTAAGTGA 2517

Scoring table: IDENTITY-NUC
Gapop 10.0, Gapext 1.0

Searched: 930621 seqs, 42862619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
N.GeneSeq-1101.*
1: /SID2/gcgdata/geneSeq/geneSeq/NA1980.DAT.*
2: /SID2/gcgdata/geneSeq/geneSeq/NA1981.DAT.*
3: /SID2/gcgdata/geneSeq/geneSeq/NA1982.DAT.*
4: /SID2/gcgdata/geneSeq/geneSeq/NA1983.DAT.*
5: /SID2/gcgdata/geneSeq/geneSeq/NA1984.DAT.*
6: /SID2/gcgdata/geneSeq/geneSeq/NA1985.DAT.*
7: /SID2/gcgdata/geneSeq/geneSeq/NA1986.DAT.*
8: /SID2/gcgdata/geneSeq/geneSeq/NA1987.DAT.*
9: /SID2/gcgdata/geneSeq/geneSeq/NA1988.DAT.*
10: /SID2/gcgdata/geneSeq/geneSeq/NA1989.DAT.*
11: /SID2/gcgdata/geneSeq/geneSeq/NA1990.DAT.*
12: /SID2/gcgdata/geneSeq/geneSeq/NA1991.DAT.*
13: /SID2/gcgdata/geneSeq/geneSeq/NA1992.DAT.*
14: /SID2/gcgdata/geneSeq/geneSeq/NA1993.DAT.*
15: /SID2/gcgdata/geneSeq/geneSeq/NA1994.DAT.*
16: /SID2/gcgdata/geneSeq/geneSeq/NA1995.DAT.*
17: /SID2/gcgdata/geneSeq/geneSeq/NA1996.DAT.*
18: /SID2/gcgdata/geneSeq/geneSeq/NA1997.DAT.*
19: /SID2/gcgdata/geneSeq/geneSeq/NA1998.DAT.*
20: /SID2/gcgdata/geneSeq/geneSeq/NA1999.DAT.*
21: /SID2/gcgdata/geneSeq/geneSeq/NA2000.DAT.*
22: /SID2/gcgdata/geneSeq/geneSeq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2517	100.0	2517	19	AAV22250
2	2506	99.6	2569	20	AAK61238
3	1041.4	41.4	2690	20	AAZ20209
4	979.6	38.9	2498	19	AAV40801
5	979	38.8	2498	22	AAZ89523
6	976.8	38.8	2690	20	AAZ20208
7	968.6	38.5	2497	20	AAZ10002
8	966.8	38.4	2343	22	AAH27438
9	956.6	38.0	2746	19	AAV40800
10	956.6	38.0	2746	20	AAZ10001
11	948.6	37.7	2524	21	AAZ00335

12	886.8	35.2	1762	19	AAV40802
13	886.8	35.2	1762	20	AAZ10003
14	868.4	34.5	1762	20	AAZ20210
15	303.8	12.1	928	20	AAZ20207
16	275.8	11.0	2780	20	AAK61239
17	243.2	9.7	2868	21	AAZ00332
18	240.8	9.6	2842	21	AAZ00334
19	220.8	8.8	2653	21	AAZ00338
20	191.8	7.6	2668	21	AAZ00337
21	184	7.3	3060	21	AAZ00333
22	174.4	6.9	1816	21	AAZ00331
23	169.6	6.7	993	19	AAV40803
24	169.6	6.7	993	20	AAZ10004
25	114	4.5	822	21	AAZ42001
26	90.8	3.6	253	20	AAK61259
27	78	3.1	936	22	AAZ58252
28	78	3.1	936	22	AAZ58254
29	78	3.1	936	22	AAZ58257
30	78	3.1	936	22	AAZ58259
31	78	3.1	936	22	AAZ58262
32	78	3.1	936	22	AAZ58255
33	77.2	3.1	936	22	AAZ58252
34	77.2	3.1	936	22	AAZ58254
35	77.2	3.1	936	22	AAZ58257
36	77.2	3.1	936	22	AAZ58259
37	77.2	3.1	936	22	AAZ58262
38	77.2	3.1	936	22	AAZ58255
39	50.4	2.0	636	22	AAH87794
40	48.2	1.9	540	22	AAH87795
41	40.4	1.6	1631	22	AAH75403
42	39.2	1.6	3408	22	AAH18536
43	39.2	1.6	66566	21	AAZ53450
44	38.8	1.5	244	22	AAZ58238
45	38.8	1.5	4590	22	AAH24065

ALIGNMENTS

RESULT 1	
ID	AAV22250 standard; CDNA to mRNA: 2517 BP.
AC	AAV22250;
DF	06-JUL-1998 (first entry)
XX	
DE	Cucurbit raffinose synthase cDNA.
XX	
KW	Cucurbit; raffinose synthase; sucrose; galactinol; ds.
OS	Cucumis sativus.
XX	
FN	Key
FT	CDS
FT	Location/Qualifiers
FT	/*tag= a
FT	/product= raffinose_synthase
PN	JPI0084973-A.
XX	
PD	07-APR-1998.
XX	
PF	28-APR-1997; 97JP-011124.
XX	
PR	26-JUL-1996; 96JP-0198079.
XX	
PR	26-APR-1996; 96JP-0107682.
XX	
PA	(AJIN) AJINOMOTO KK.
XX	
DR	WPI: 1998-264858/24.
XX	
PT	P-PSDB; AAW53570.
XX	
PT	Raffinose synthase gene - useful for preparation of raffinose in

transformed plant

Claim 7: Pages 17-20; 26pp; Japanese.

The present sequence encodes cucumber raffinose synthase, which forms raffinose from sucrose and galactinol, has an optimum pH of 6 to 8 and working temperature of 35 to 40 degrees C, has a molecular weight of 75 to 95 kDa by gel filtration or 90 to 100 kDa by PAGE and SDS-PAGE under reductive conditions and is inhibited by lodosacetamide, N-ethylmaleimide and myoinositol.

Sequence 2517 BP; 613 A; 524 C; 642 G; 738 T; 0 other;

Query Match 100.0%; Score 2517; DB 19; Length 2517;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2517; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 AAAAACAACCCCTTCTTTAGTTTGGTTTGGTTTCTTTCTTTCTTCAACAAATGGC 60
1 aaaaaaacacccttctttagtttgggttcttcttcttcttcttcttcaacaatggc 60
61 TCCAGCTTTAAAAATGTCGCTCCACAGTAGTTTCATTGGATGGCTTAATGACATGTC 120
61 tccagcttttaaaaatgtctcctccacagtagtttcaatcgaagcttaaatgacatgtc 120
121 GTCACGCTTTGCAATCGACGAGTGGATTTCATCTGTACAGCGTTCATTCTGTCCGA 180
121 gtccagctttgcaatcgacgagtggtatttcacatcgtacagcttcattctgtccga 180
181 TGTTCCTGAGAACATTGTGCTTCTCTCTCCGTACACTTCGATPAGAACAGTCCCGGT 240
181 tgttcctgagaaactgtctctctctctcgtacacttcgaatagaacaggtcccggt 240
241 TTGGGTGGTGGCTTGTGGATTGCAAGCGCTGGAACTGATGACCGACATGTTGTTTC 300
241 ttgggtgggtgtgtgtggattgcagcgctggaaactgatgacccgacatgtttgtttc 300
301 GATTGGGAAGCTGGAAGATATTCGGTTATGAGTATTTTACGTTTAAAGTTTGGTGAC 360
301 gattgggaagctggaagatatttcggttatgagtattttacggtttaaagtttggtgac 360
361 TACACACTGGGTGGTGGAAATGTCGGGATCTTGAATCGAGACTCAGATTGTGATCCT 420
361 tacacactgggtgtgtggaaatgtcgggatcttgaatcgagactcagattgtgactct 420
421 TGGAAGTCGAGATTCTGGTGCAGCGTATGTTTCTTCCGATCGTTGAGGACCGCTT 480
421 tggaaagtcgagattctgggtgcagcgatgtttcttccgatcgttggagggacgctt 480
481 CCGAAGCTCGATTGAGCTGGGATGATGACTTGTGCGATGTTGTGTGAGAGTGGTTC 540
481 ccgaagctcgattgagctgggagtgatgacttgtgcgattgttgtgtgagagtgttc 540
541 GTCAAAAGTTTGTATGATGATGTTCCGAAGTATGTTGATCTTATGCTGGTATGATCC 600
541 gtcaaaagtttgtatgatgatgttccgaagtattgttgatcttattgctgtgatgacc 600
601 GTTTCACATTGTTAAAGGCGATGAAGATGTGAGACCCATCTTGAAATTTTTCGCTT 660
601 gtttcacattgtttaaaggcgatgaagatgtgagaccccatcttgaaatttttcgctt 660
661 GTTGGAGAGAAGATCTCACAGGATATCGTGACAAATTCGGTTGGTGACGTGGGACGC 720
661 gttggagagaagatctcacaggatattcggtgacaattcggttggtgacgtgggacgc 720
721 GTTTTAACTTAACGGTTTCAACAGGCGTAAATAGAAAGCGTAGAGCATCTCTGACAGG 780
721 gtttttaacttaacggtttcaacaggcgtaaatagaaagcgtagagcatctctgacagg 780
781 CGGTTGTCCTCCGGTTTAACTTAATGAGAGATGTTGGCAATCATCGAGACAGATTTC 840
781 cggttgtctcccggtttaaacttaatgagagatgttggcaatcatcgagacagatttc 840

841 GGATCCCATCACCAAGAAGATGAACCAACCGTCGCCGGGACCAATGCCCTGCCG 900
841 ggaatcccatcaccaagaagatgaaccaacccgtccggggacccaatgccctgccg 900
901 TCTTTTGAATTTCCAAAGAGATTACAAATTCGCGACTACGTCATATCCCAAGCCACCG 960
901 tcttttgaattttccaaagagattacaaattcgcgactacgtcataatcccaagccacg 960
961 CCCCCAGCCCGCCAGAGAGGATGAAGCGCTTTATGATGAATCAAGAGAGATTAA 1020
961 cccccagcccgccagagaggatgaagcgctttatgatgaatcaagagagatttaa 1020
1021 GACTGTCGACATGTTTATGTTTGGCATGCTTGTGTGATATTTGGGTCGCTTGGCCC 1080
1021 gactgtcgcacatgtttatgtttggcatgcttgtgtgatattttgggtcgcttggccc 1080
1081 GCAGGTCCTGGCTTGGCTGAGGACGTGATTCAGCCAGTGGCTTACACAGGCTGCA 1140
1081 gcaggtcctggcttggctgaggacgtgatccagcttgcacaggttcacaggtcgca 1140
1141 GATGACGATGAGAGATTGGCGGTGATGAAGATTGTTCTTATAGGTGGCGCTGTCC 1200
1141 gatgacgatgagagattggcgggtgatgaagattgttcttataggtggcgctgtcc 1200
1201 GCCCGAAGAGCTTACGAGATGTACGAAGACCTTCATGCTTATGGAAAAGTTGGGAT 1260
1201 gcccgaaagagcttacgagatgtacgaagaccttcacatcttgaaaaagttgggat 1260
1261 CGACGGTGTAAAGATTGACGTTATCCACTATTGAGATGTTGTGTAAGACTATGAGG 1320
1261 cgacgggtgtaaagattgacgttatccactattgagatgttgtgtaagactatgagg 1320
1321 GAGAGTGATTTGGCAAAAGCATATTACAAAGCAATGACCAATCAATAATAACATTT 1380
1321 gagagtgatttggcaaaagcatatttacaaagcaatgacccaataataataaacatt 1380
1381 TAAAGGAATGAGATCTTCCATGCAATGGAACATTTGACACTTCACTTCTTGGCAC 1440
1381 taaaggaatgagatcttccatgcaatggaaactttgacacttcaacttcttggcac 1440
1441 GGAAGCTATCTCTTGTGCTGCTGTTGATGACTTGTGTGACGAGACCCCTCTGTGA 1500
1441 ggaagctatctcttgtgctgctgttgatgacttgtgtgacgagacccctctgtga 1500
1501 TCCAAACGATGTTTGGCTCCAAAGATGTCACATGCTTATGTCGCAACGACACTT 1560
1501 tccaaacgatgttggctccaaagatgtcacatgcttattgtgtccaaagacactt 1560
1561 GTGATGCGGAACTTCATCCACCTGACTGGGATATGTTCCAAATCCACCCCTTGTGC 1620
1561 gtgatgcggaacttcattccacctgactgggatattgttccaaatccaccccttgtgc 1620
1621 CGCCTTCATGCTGCTCTGAGCCATCTGTGTCGCCGATCTATGTAGTGAATTCGT 1680
1621 cgccttcacatgctgctctgagccatctgtgtgcccgatctatgttagtgaattcgt 1680
1681 GGAAGACATATCTTGAATCTTTCGAAAAAATCTAGTCTTCTTATGATGATCCTTTCG 1740
1681 ggaagacatattctgaatcttctcgaaaaaactagtccttcctatgatatgactctcg 1740
1741 AAGTGAGTACATGACCTCCGACGCGATGTTGTTGTAACACCTTGCATTAATGG 1800
1741 aagtgaagtacatgacctccgacgcatgttgttggtaaacaccttgcattatgg 1800
1801 AAGTGAAGTACATGACCTCCGACGCGATGTTGTTGTAACACCTTGCATTAATGG 1860
1801 aagtgaagtacatgacctccgacgcatgttgttggtaaacaccttgcattatgg 1860
1861 CTGCCAAGAGAGATGCTGTCGAGACACGCGCAACCAATGCTTTTACATATCTC 1920
1861 ctgccaaagagagatgctgtcgaacacgcgcaaccaatgcttttacaatactc 1920

QY 1921 AAAAGAGTGCATCCAAACTACCCAAAGACATAGATGGCACAGTGAGAAAACCC 1980
DB 1921 AAAAGAGTGCATCCAAACTACCCAAAGACATAGATGGCACAGTGAGAAAACCC 1980
QY 1981 TATCTCTATGAGGCGTTAAACCTTGGCTTTACTCTATGACCAAAACCTTAT 2040
DB 1981 TATCTCTATGAGGCGTTAAACCTTGGCTTTACTCTATGACCAAAACCTTAT 2040
QY 2041 CCTTCGCAAGCCCTCCAAAGATCTGACATAGCTTGGCCATTCGAATTCGACTCAT 2100
DB 2041 CCTTCGCAAGCCCTCCAAAGATCTGACATAGCTTGGCCATTCGAATTCGACTCAT 2100
QY 2101 CACTGTTTCACGAGTACCAAACTCATCAACTCTGACACTTTGGCCAAATGGCGT 2160
DB 2101 CACTGTTTCACGAGTACCAAACTCATCAACTCTGACACTTTGGCCAAATGGCGT 2160
QY 2161 GGTGACATGCTTAACTAGTGAAGCCATTCGATCTGACATAGACATGACCTAAG 2220
DB 2161 GGTGACATGCTTAACTAGTGAAGCCATTCGATCTGACATAGACATGACCTAAG 2220
QY 2221 CTCGATCGAGATTTGGTCAAGGTCGTCGATCGAGATTTGGTCAAGGTCGTCGAT 2280
DB 2221 CTCGATCGAGATTTGGTCAAGGTCGTCGATCGAGATTTGGTCAAGGTCGTCGAT 2280
QY 2281 AAGGCTTGTCTGATTTGATGAGGAGATTTGGTCAAGGTCGATGATGATGATGATGAT 2340
DB 2281 AAGGCTTGTCTGATTTGATGAGGAGATTTGGTCAAGGTCGATGATGATGATGATGAT 2340
QY 2341 GGTGCTTCAAGTTCGATGAGGAGATTTGGTCAAGGTCGATGATGATGATGATGATGAT 2400
DB 2341 GGTGCTTCAAGTTCGATGAGGAGATTTGGTCAAGGTCGATGATGATGATGATGATGAT 2400
QY 2401 CTTGTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 2460
DB 2401 CTTGTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 2460
QY 2461 ATCAATGATTTCTCTCCAAAGAAATATGATGATTTGAGAGATATTAATGA 2517
DB 2461 ATCAATGATTTCTCTCCAAAGAAATATGATGATTTGAGAGATATTAATGA 2517

RESULT 2
AAK61238
ID AAK61238 standard: cDNA to mRNA; 2569 BP.
XX AAK61238;
AC
XX 29-JUL-1999 (first entry)
DT
XX
DE Cucumber raffinose synthase encoding cDNA.
XX
KM Raffinose synthase; sucrose; galactinol; ds.
OS Cucumis sativus.
XX
OS
FH Key Location/Qualifiers
FT CDS 56..2410
ET /-tag- a
XX
XX JP11123080-A.
PN
XX
PD 11-MAY-1999.
XX
XX
PF 24-OCT-1997; 97JP-0292969.
PR
XX 24-OCT-1997; 97JP-0292969.
XX
PA (AJIN) AJINOMOTO KK.
XX
XX WPI; 1999-340516/29.
XX
XX P-PSDB; AA117417.
XX
XX New raffinose synthase gene - for production of raffinose from

PT sucrose and galactinol
XX
XX
PS Claim 10; Page 22-25; 37pp; Japanese.
XX
CC The present invention describes a raffinose synthase, having an activity
CC of forming raffinose from sucrose and galactinol. The raffinose synthase
CC gene can be used for expression in a plant for the production of
CC raffinose. The raffinose synthase can give raffinose from sucrose
CC and galactinol efficiently. The present sequence encodes raffinose
CC synthase from cucumber.
XX
XX
SQ Sequence 2569 BP; 641 A; 527 C; 643 G; 755 T; 3 other:

Query Match 99.6%; Score 2506; DB 20; Length 2569;
Best Local Similarity 100.0%; Pred. NO. 0;
Matches 2517; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 AAAAAAAAAAAGGCTTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 60
DB 1 AAAAAAAAAAAGGCTTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 60
QY 61 TCCATGTTTAAAAATGCTGCTCCAGCTAGCTTCAATTTGATGCTTAAATGACATGTC 120
DB 61 TCCATGTTTAAAAATGCTGCTCCAGCTAGCTTCAATTTGATGCTTAAATGACATGTC 120
QY 121 CTCACCGTTTGAATTCGACGATTCGATTTCACTGTAACGCTGCAATTCGTTCTGTCGCA 180
DB 121 CTCACCGTTTGAATTCGACGATTCGATTTCACTGTAACGCTGCAATTCGTTCTGTCGCA 180
QY 181 TGTTCCTGAGACATTTGTTGCTTCTCTCTCCGATACACTTGTGATAGACAAATCCCGGT 240
DB 181 TGTTCCTGAGACATTTGTTGCTTCTCTCTCCGATACACTTGTGATAGACAAATCCCGGT 240
QY 241 TTTGCTGCTGCTTGTGATTTGATTCGACGCTGCGAACCCTGATAGCCGATGTTGTTTC 300
DB 241 TTTGCTGCTGCTTGTGATTTGATTCGACGCTGCGAACCCTGATAGCCGATGTTGTTTC 300
QY 241 TTTGCTGCTGCTTGTGATTTGATTCGACGCTGCGAACCCTGATAGCCGATGTTGTTTC 300
DB 241 TTTGCTGCTGCTTGTGATTTGATTCGACGCTGCGAACCCTGATAGCCGATGTTGTTTC 300
QY 301 GATTTGGGAAGCTGGAAGATTCGCTTTTATGATTTTCAAGTTTAAAGTTTGGTGGGTC 360
DB 301 GATTTGGGAAGCTGGAAGATTCGCTTTTATGATTTTCAAGTTTAAAGTTTGGTGGGTC 360
QY 361 TACACACTGGTGTGTCGAATGTCGAGGATTCGATGATGAGACATGATGTCCT 420
DB 361 TACACACTGGTGTGTCGAATGTCGAGGATTCGATGATGAGACATGATGTCCT 420
QY 421 TGAGAAGTCAGATTTCTGTCGACCGTATGTTTCTTCTTCCGATTCGTGAGGACCGTT 480
DB 421 TGAGAAGTCAGATTTCTGTCGACCGTATGTTTCTTCTTCCGATTCGTGAGGACCGTT 480
QY 481 CCGAACCCTGATTCAGCCCTGGGAGATGATGATTTGTCGATGTTTGTGTCGAGAGTTC 540
DB 481 CCGAACCCTGATTCAGCCCTGGGAGATGATGATTTGTCGATGTTTGTGTCGAGAGTTC 540
QY 541 GTGGAAGTTGTTGATGATTCGCTTCGGAAGTATGTTGATTTCTTCAATGCTGATGATGTC 600
DB 541 GTGGAAGTTGTTGATGATTCGCTTCGGAAGTATGTTGATTTCTTCAATGCTGATGATGTC 600
QY 601 GTTTCACCTGTTTAAAGAGCGATGAAGATTCGTGAGACCCATCTTGAACCTTTTCGCTT 660
DB 601 GTTTCACCTGTTTAAAGAGCGATGAAGATTCGTGAGACCCATCTTGAACCTTTTCGCTT 660
QY 661 GTTGGAGGAGAGACTCCACAGATTCCTGTGACAAATTCGTTGGTGTGACATGGGACGC 720
DB 661 GTTGGAGGAGAGACTCCACAGATTCCTGTGACAAATTCGTTGGTGTGACATGGGACGC 720
QY 721 GTTTTACCTAAGGTTCAATCCACAGGCTAATTAAGAGGCTGAGGCACTGTCGACGG 780
DB 721 GTTTTACCTAAGGTTCAATCCACAGGCTAATTAAGAGGCTGAGGCACTGTCGACGG 780
QY 781 CGGTTGCTCCCGGTTTATGCTTAATGACAGATGGTTGGCAATCCATCGACACGATTC 840
DB 781 CGGTTGCTCCCGGTTTATGCTTAATGACAGATGGTTGGCAATCCATCGACACGATTC 840

QY 841 GGATCCATTCACCAAGAAATGAACCAACCGTCGCCGAGCAAAATGCCCTGCCG 900
|||||
Db 841 ggtcccatcaccaagaagaaatgaaccaaaccgctcgcgagcaaatgcccgcg 900
QY 901 TCTTTTGAATTCAGAGAGATTAACAATTCGCTGACTAGCTCAATCCCAAGGCGACCGG 960
|||||
Db 901 tcttttgaattccaagaagaaatgaacaaatccgtgactaacgtaacccaagccacccg 960
QY 961 CCCCGAGCGCGCAGAAAGGAGATGAAGCGCTTTATAGATGAATCAAGAGAGATTAA 1020
|||||
Db 961 ccccgagccgcgccagaaagagatgaagcgcttataagatgaactcaaaagagatttaa 1020
QY 1021 GACTGTGAGACATGTTTATGTTTGGCATGCTTGTGTGATTTGGGGTGCCTTGCCCC 1080
|||||
Db 1021 gactgtgagacatgtttatgttttgacatgtcttgtagatattg99gtgcccctgcgccc 1080
QY 1081 GCAGGTGCTGGCTTGCTTACAGCAGCTGTGATTCAGCCAGCTGCTTACACAGGCTGCA 1140
|||||
Db 1081 gcaggtgctgcttgcttgcctgagcaagctgtgactcagccagctcttcacaa9gctgca 1140
QY 1141 GATGAGATGAGAGATTGGCGGTGATAGATTGTTCTTCATTAAGTTCGGGCTGCTCCC 1200
|||||
Db 1141 gatgagatgagagattggcgggtgataagattgtcttcataaagtc9gctgctccc 1200
QY 1201 GCCGGAAGAGCTGAGAGAGATGTACGAAGGACTTCATGCTCATTTGGAAAAAGTTGGGAT 1260
|||||
Db 1201 gccggaagagctgagagagatgtacgaaggacttcacgtccatttggaaaaagtlggat 1260
QY 1261 CGACGCTGTTTAAGATTGACCTTTATCCACCTATTGGAGATTTGTGTGAACATRTGAGG 1320
|||||
Db 1261 cgacgctgtttaagattgacgtttatccacctatgagagatgtgtgtaaaactatlgaa9 1320
QY 1321 GAGAGTGATTTGGCAAGGCATATTACAAAGCATGACCAAAATCAATAATAAATTTT 1380
|||||
Db 1321 gagagtgatlttgcaaggcataattacaagcaatgaaccaataataaactltt 1380
QY 1381 TAAAGAAATGAGATCAATTCGAAGTATGAAACATTTGTAAGACTTCATGTTCTTGGCAC 1440
|||||
Db 1381 taagaaatgagatcatttcgaagcaatltgaacgaacttcattgctcctlgcac 1440
QY 1441 GGAAGCTATCTCTCTGTGTGTGTGTGTGATGACTTTGGTGCAGCAACCCCTGTGTGA 1500
|||||
Db 1441 ggaagctatctctctgtgtgtgtgtgtgactttgtgtgacacccctctgtgtga 1500
QY 1501 TTCAAACGATGACGTTTGGCTCCAAAGATGTCACATGTTGATGTGCCAAGCAGACGCTT 1560
|||||
Db 1501 ttcaaacgatgacgttttggctccaaagatgtcacatgttcaatgtgtgccaa9acgctt 1560
QY 1561 GTGATGGGGAACCTTCATCCACCCCTGACTGGGATATGTTCCAAATCCACCAACCCCTGTGC 1620
|||||
Db 1561 gtgatggggaaccttcacccctgactgggatatgttccaatcccaacccctltgac 1620
QY 1621 CGCCTTCATGCTGCTCTGTGAGCCATCTCTGTGGGCCCATCTATGTTAGTATTCTGT 1680
|||||
Db 1621 cgcttccatgctgctctgtgagccatctctgtggtgcccgaatctatgttagttatctgt 1680
QY 1681 GGGAAACATTAACCTTGATGCTGTGAAAAAATCTAGTCTTCATGATGATGATGATCTTCG 1740
|||||
Db 1681 gggaaacataacttctgacttctgaaaaaactagtgcttccctgaatgagatcccttcg 1740
QY 1741 AAGTAGTACTATGCACTCCGACTCGCATGTTGTTTGTGAAAGACCTTTGCATTAATGG 1800
|||||
Db 1741 aagtagtactatgcactccgactcgcatgtgtgttgaagaaccttltgcataatg 1800
QY 1801 AGAAGCATATGCTTAAAGTTTGGAAATCTCAACAAGTTCACATCTACTCTGATGATTCAA 1860
|||||
Db 1801 agaagcatatgctttaaagtttggaaatctcaacaagttcactctgagatgtatgtgtcattcaa 1860
QY 1861 CTGCGAAGAGGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1920
|||||
Db 1861 ctgcgaaagagagagatgagatgagatgagatgagatgagatgagatgagatgagatgagatg 1920

QY 1921 AAAACGATGACATCCAAAATAACCCAAAAGACATTAGATTGCGACAGTGGAGAAAACC 1980
|||||
Db 1921 aaaaagatgacatccaaaataacccaaaagacatagaaatgacagatgag9a9aaacc 1980
QY 1981 TATCTCTATTGAAGGCGTTTAAACCTTTGGCTTTACCTCTATCAAGCCAAAAACTTAT 2040
|||||
Db 1981 tatctctattgaagcgctttaaacccttgccttaccctctatccaagccaaaaacttat 2040
QY 2041 CCTCTCAAGCCCTCTCAAGATCTTGACATAGCTCTTGACACCTTTGGAAATTCAGAGTCAT 2100
|||||
Db 2041 cctctcaagccctctcaagatcttgacatagctctgacccaatcgaatlc9agtcac 2100
QY 2101 CACTGTTTCCACAGTGACCAAACTCAACCACTTCCTACACTTTGGCCCAATTGGGCT 2160
|||||
Db 2101 cactgtttccacagtgaccaactcaccaactctctcactcttgcacatltggcct 2160
QY 2161 GGTGAACATGCTTTAACACTAGTGAGAGCCATCAATCTGTGAGCTATGACATGACCTTAAG 2220
|||||
Db 2161 ggtgaacatgcttttaacactagtgagagccatccaatctgtgactatgagatgacctaa9 2220
QY 2221 CTCAGTCGAGATTGTGTGCAAGGGTGTGTGAGATGCGAGTATTTGCATCGAAAAAAC 2280
|||||
Db 2221 ctcagtcgagatltgtgtccaaggtgtgtgaagatgcagatlttgcatcga9aaaacc 2280
QY 2281 AAGGCTTGTGCTATTGATGATGGAGAGATGTTGGTTCAAGTATGATCAGGACCAATGAT 2340
|||||
Db 2281 aaggcttgtgctattgataatggaagatgttggttcaagatgatacga9aacatgct 2340
QY 2341 GGTGTTCAAGTGCGCATGCGCAATTTGATCTTCATCGGGTGGCATTTGGTTATCGAGTA 2400
|||||
Db 2341 ggtgttcaagtgcatgcgcaatttgatcttcacggtggcagatlttcgttalcg9tia 2400
QY 2401 CTGTTTAAATTTTATTTATTTATGTA-AGCTCAATGATTTGTTGTTTGCCTGTGTTGC 2459
|||||
Db 2401 ctgttttaattttatlttatgtatgacatgattgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2459
QY 2460 TATCAATGATTTCTCTCCAAAGAAATATGATGTTGAGATTAATTAATGTA 2517
|||||
Db 2461 tatcaatgatlttctctccaaagaaatataatgtatgttaattgtgag9aatlaa9tga 2517

RESULT 3
AAZ20209 standard; cDNA; 2690 BP.
ID AAZ20209 standard; cDNA; 2690 BP.
XX
AC AAZ20209;
XX
DT 17-JAN-2000 (first entry)
XX
DE Mustard raffinose synthase cDNA.
XX
KW Raffinose synthase; mustard; transgenic plant; ss.
XX
OS Brassica juncea.
XX
FH key Location/Qualifiers
FT CDS 134..2467
FT /*tag= a
FT /note= "this region is specifically claimed in
FT Claim 7"
XX
XX EP953643-A2.
XX
XX 03-NOV-1999.
XX
XX 27-APR-1999; 99EP-0107430.
XX
XX 30-APR-1998; 98JP-0120550.
XX
XX 30-APR-1998; 98JP-0120551.
XX
XX 04-DEC-1998; 98JP-0345590.
XX
XX 10-DEC-1998; 98JP-0351246.
XX
XX (SUMO) SUMITOMO CHEM CO LTD.

XX Watanabe E, Oeda K;
PI
XX
DR MPI: 1999-593144/51.
P-PSDB; AAY32074.
XX
PT New sense and antisense genes, useful for altering the level of
PT raffinose in food plants -
XX
PS Claim 7: Page 32-36; 55pp; English.
XX
CC This is the nucleotide sequence of a mustard cDNA clone coding
CC for raffinose synthase (see AAY32074), a protein which can bind a
CC D-galactosyl group through an alpha(1-6) bond to the hydroxy group
CC attached to the carbon atom at the 6-position of the D-glucose
CC residue in a sucrose molecule to form raffinose. The cDNA was
CC isolated from mustard (Brassica juncea) leaf cDNA by PCR. Probes
CC or primers generated from plant raffinose synthase genes (see
CC AAY3207-10) may be used to obtain other raffinose synthase genes
CC by labeled detection or amplification (claimed). These genes may
CC be used to control the levels of raffinose produced in plants.
CC Antisense genes can be used to knock out existing gene activity,
CC and sense genes to increase the level of gene activity. The
CC resulting transgenic plants may be used as a food source to alter
CC the growing conditions for gut enterobacteria, providing general
CC health advantages. The sense and antisense genes may also be used
CC in gene and phenotypic analysis of plants and for the selection of
CC plant clones with the desired characteristics with respect to
CC raffinose content.
XX
SQ Sequence 2690 BP; 648 A; 608 C; 668 G; 765 T; 1 other;

Query Match 41.4%; Score 1041.4; DB 20; Length 2690;
Best Local Similarity 68.3%; Pred. No. 0;
Matches 1530; Conservative 0; Mismatches 676; Indels 33; Gaps 5;

OY 135 TCAGCGATGGAGTTTCACGTGTAAGCGTCATTCGTTGTCGGATTCCTGAGAC 194
DB 210 tgaaggcttcgcagctcctcagtcgaagtcagctgctcctcaacgcgagttacg 269
OY 195 TTGTTGCTTCCTCTCTCCCTAC-----ACTTCGATAGCAAGTCCCGGTTTCG 244
DB 270 tgaactgactgctcctcctcctcctcagtcgaagcagcgagcagcgcctcgc 329
OY 245 GTTGTGCTTCTTGTGATTCGACGCGTCG--GAACTGATAGCCGACATGTTTTCGA 302
DB 330 ctggttcattcattcgggttcaatcctcgaagtcgagccagcagcgctgctcca 389
OY 303 TTGGGAGCTGAAGATATTCGGTTTATGATTTTTCAGGTTTAACTTGGTGACTA 362
DB 390 tcggttaaaccaabggatcgcattcagatcagacatccgctcctcaagcttgcggacta 449
OY 363 CACACTGGTGTGTCGAATGATGGGATCTTGAATCGAGACTCAAGATTGATCTTG 422
DB 450 ctccatcggtcgtcttcaaaagatcgaacatcgagaaagccagatcattcctcg 509
OY 423 AGAAGTACAGATTTCGTGTCGACCGATATTTCTTCTTCGATCGTGAAGGACCTTCC 482
DB 510 agaactcggggtcggtcctcattatgctctcctcgcgcgtcttgaagagctcttc 569
OY 483 GAACCTGATTCAGCCCTGGGATGATGATCTTTCGATGTTTGTGAGAGTGTTTGGT 542
DB 570 gtccatcctctcagccctggggaagacgaatgacgtgcggttctgttcgaaatccgggtcga 629
OY 543 CGAAAGTTGTGATGATCGTTCGAAATGTTGTTATCTTCATCTGATGATGATCCGT 602
DB 630 cccaggtgacccgggtcggagttctcgcaagttgtagtctcaagccggagacatccgt 689
OY 603 TTGCACTTTTAAAGAGCCATGAAGATCGTGAAGACCCATCTTGGAACTTTTCGCTTGT 662
DB 690 tcaagctcgtgaagacgactgaaagtggttaggtgctcattatgaacacatcctcaagctct 749

OY 663 TGAGAGAGAAGACTCCAGCAGGATATCGTGACAAATTCGGTTGGTCACGTGGACCGCT 722
DB 750 tggaaagagaagacrtccgcgggaaatcgtcgaatgaagttcgggtgtagcgtggagctgc 809
OY 723 TTTACTTAACGGTTTCATCCACAGGGCGTAAATAGAACGCTGAGGCACTTCGTGACGGCG 782
DB 810 ttctattgacggtgaaccccgagcagagttcaataaggtgttaagttctcgtcgaaggtg 869
OY 783 GTTGTCTCCCGGTTTAACTGATGATGATGATGATGATGATGATGATGATGATGATG 842
DB 870 gtgtccgcgggagtgtgtcctcaatcgaacgaagtcgtggaatcgtatgtaacatccgcg 929
OY 843 ATCCATTCACAAAGAGAGATGACCAACCCGTCCGCGGAGCAAAATCCCTGCGCTC 902
DB 930 atgtatcagatgttgaagagatgagttgacgcgtccgcggggacaaatgcttctgaagc 989
OY 903 TTTTGAATTTCCAAAGAAATTCAAATTCCTGACATGCTCAATCCCAAGGCCACCGGCC 962
DB 990 ttctgaatttcagaagaaactcaagttcagagactacgtctcctcgaagaaacaaaa--- 1046
OY 963 CCGAGCCGCGCCAGAGGGGATGAGGCGCTTATGATGAACTCAAGAGAGACTTTAAGA 1022
DB 1047 -----agaaagtcgggaatgaagcttcgtcgaagatcgaagaagaatctcca 1097
OY 1023 CTGTGAGCATGTTTAAATGTTGGCATCTTGTGTGATATTGGGGTGCGCTTCGCCCGC 1082
DB 1098 cgtgtatcattacatcattcgtcgaacgcttcgtcgaacgttggtgtcttcgcgcgc 1157
OY 1083 AGGTGCTGCTGCTTCCTGAGGACGCTGTGATTCAGCCAGTGTCTTACCAAGGCTCAG 1142
DB 1158 gagctcctactcttcgcgcctcctcaactatgttcgcgaagagctcgcgggcttaag 1217
OY 1143 TGACATGAGAGATTTTGGCGGTGATTAAGATTTGTTCTTCTTAAGTCCGGCTGCTCCGC 1202
DB 1218 tgaagatgcaagatctcgcgttgaataagatgtcgaatcaggaatcggatctgcgcgc 1277
OY 1203 CGAGAGAGCGCTGAGAGATGATGACGAAGACTTCATCTCTTGTGAAAAGTTGGGATCG 1262
DB 1278 cggaaatcgggaaatgagtttcaagaagttctcaacccctcctcaaaagaatcgtgata 1337
OY 1263 ACGGTGTTAAGATTTGACGTTATCCACCTATTTGAGATGTTGTGTAAGACTATGAGGGA 1322
DB 1338 acggtttaaagttcagtcgtatccacatattgagatgtgtcgggaaatcagcgagg 1397
OY 1323 GATGCGATTTGGCAAGAGCATATTACAAAGCAATGACCAATCAATTAATTAATTTTA 1382
DB 1398 gagtagagcttggttaaggtcttactcaagagcgttaactctcctcaatgaagcatttg 1457
OY 1383 AAGGAAATGAGATTCATTCGAATGATGAAATGTAAGCACTTCATGTTCTTGCGACGG 1442
DB 1458 acggttaacggttaltcgtcagatcagatggaacatgaaatgattcattccttggaacg 1517
OY 1443 AAGCTATCTCTTGTGCTGTTGTGTGATGATCTTTGTGTCACGACCCCTGTGTGATC 1502
DB 1518 aagcatcctctcagtgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1577
OY 1503 CAACGCTAGCTTTGGCTGCAAGATGCAATGATGATGATGATGATGATGATGATGATG 1562
DB 1578 taaagcgaacatattggtcgtgaagaaatgcaacatggtccacgtgctcctcaaaagctct 1637
OY 1563 GGAATGGGAACTTCATCCACCTGACTGAGATGATGATGATGATGATGATGATGATGATG 1622
DB 1638 ggaatggaaatcttcaatccacccgtatggaacatggttcaagtcacacatccttgctg 1697
OY 1623 CTTTCATGCTGCTTCGACGCACTCTGTGTCGCCGATGATGATGATGATGATGATGATG 1682
DB 1698 agttccatcgtctcctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 1757
OY 1683 GAAAGCATTAATCTTGAATCTTGAATAAAACTAGAGCTCTGATGATGATGATGATGATG 1742
DB 1758 gccagcagatcctcgaactccttgaagacactcgtctcgtcgtcgtcgtcgtcgtcgt 1817
OY 1743 GTGATGACTATGCACTCCGACCTCGGATGTTGTTTGAAGACCTTTTCATTAATGAGAG 1802

[illegible]

Query Match	38.9%; Score 979.6; DB 19; Length 2498;
Best Local Similarity	65.2%; Pred. No. 4.8e-286;
Matches 1536; Conservative	0; Mismatches 785; Indels 36; Gaps 5;
18-DEC-1997; 97EP-0122417.	
18-DEC-1996; 96JP-0338673.	
(SUMO) SUMITOMO CHEM CO LTD.	
Oeda K, Wantanabe E;	
WPI: 1998-324670/29.	
P-PSDB; AAW57887.	
New nucleic acid molecule encoding plant raffinose synthetase -	
capable of producing raffinose, used as food additives with	
beneficial effects on gastrointestinal flora	
Claim 1; Page 31-34; 44pp; English.	
This sequence encodes the soybean raffinose synthetase of the	
invention. The raffinose synthetase is capable of producing raffinose by	
combining a D-galactosyl group through an alpha (1-6) bond with a	
hydroxyl group attached to the carbon atom at position 6 of a D-glucose	
residue in a sucrose molecule. The DNA can be used to modify metabolism	
of a host organism by introducing into the host organism or its cell so	
that the content of the raffinose family oligosaccharides in the host	
organism or cell is changed. Raffinose oligosaccharides are useful as	
food additives with beneficial effects on the gastrointestinal flora.	
Sequence 2498 BP; 650 A; 521 C; 673 G; 653 T; 1 other;	
39 TTTCTTTCTTCACAAATGCTCTCCTAGTTTAAANAATGCTGGCTCCAAACGTAGTTTCAT	98
111 111 111 1111111 11 11 11	
45 tcttgaattcagtcoccatgctccaagaacataagaacaaacttggaaactaatcaattcgt	104
99 TTGATGCGCTTAATGACATGCTGCTCACCGCTTTCGAATGACGAGATCGGATTTTCACGTGA	158
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111	
105 gctctgtcaacaggttaatttgccttgttccataaccctlaagaagataaatttccctcgca	164
159 ACGGTCATTGCTTTGTCGTCGATGTTCTCGAACAACATTTGTTGCTTCTCTCCGTAC	217
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111	
165 acggccaccccttctcacaagaaagttccgaaaataatagtaaccccttaccatcgcg	224
218 -----ACTTCGATAGACAACTCCCGGTTTCGGTGTGTTGCTTGTGATTTGG	266
11 11	
225 acgcacaagatagtaagaacaagaaagaaagaaagcgtcgtatgtctcgtggttc	284
267 ACGCGTCGGAACCTGATPACCGACATGTTGTTGATTTGGTGAAGCGAAGATATTCGCT	326
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111	
285 acgcagcagcagccagaagcgcagcagcgttccctccgctggaagctcgaagaaataa	344
327 TTATGAGTATTTTACAGTTTAAAGTTGGTGGACTACACACTGGGTTGGTCGAATGTG	386
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111	
345 tcaatgacataatccggtttaagtggtgttgtaaccactcaacgtgcgttagcaacgagc	404
387 GGGATCTGAATCGAGACTCAGATTGTGATCCTTGAGAAAGTCAG--ATTCTGCTGAC	443
11 11 11 111 111 111 111 11 11 11 11 11 11 11 11 11 11 11 11 11 11	
405 acgaactagcagcagcagacacagatgatcttctcgaacaaaacgaccagctcgcgacgc	464
444 CGTATGTTTCTCTCTTCGCGATCGTTGAGGAGACCGTTTCGGAACCTCGAATTCAGCCCGGG	503
11 11	
465 ccttctgttgaattcccgatccctcacaagcctcgttccgagcctccctcgacaacccggtt	524
504 ATGATGACTTTGCGATGTTTGGTGGCAAGAGTCTCGTCGCAAAAGTGTGATGATCGT	563
111 111 11 11 111 111 111 111 11 11 11 11 11 11 11 11 11 11 11 11 11	
525 tgaatgataatcagtggaacgtttgcatatggaagcgggttcgacaacgtgtctgtgctccaagct	584
564 TCGGAGTATGTTGTTTCCTTCATGCTGGTATGATCCGTTTGCACTGTTTAAAGAGCGCA	623
11 11	
585 tcggagagctgcgtataatcagtcacagcttggcacaatgacccgtatcagttcgttaagaaagca	644


```

OY 1881 GTCGTAGACGACCGCAACCAATGCTTTTACATACTCAAAAGACGATCCAAA 1940
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1890 gtccgtaactgaggaagaagaagatgcctccgaattctcaaaactgagacgtctag 1949
OY 1941 CTAAACCAAAAGACATAGATGGACAGTAGAGAAAACCTATCTATATGAAGCGTTA 2000
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1950 cgagtcctcaagaacattgaatgagcaatgggaaagcccaatgatcatlaaaggatga 2009
OY 2001 AAGCTTTGGCTTTTACCTATCAACCAAAACCTATCTCTCCCAAGCCCTCTCAAG 2060
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2010 atgtgtctgtctatattgttcaagaccacaactaaagatcatgaagcatagaga 2069
OY 2061 ATCTTGACATACCTCTGTGACCTGCAATTCGACTCATCTGTTCCACAGTAGACA 2120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2070 aattggaagttcacctgagccattacttctgagctatctgacagtgctcccaagtctg 2129
OY 2121 AACCTATCCAAACTTCTTACACTTTGCCCAATTTGGGCTGTGAACATGCTTAACACTA 2180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2130 tgcgtcaaaaaaagttaattcaattgtcccaattgtagtgaatgtaacgtctaaacatg 2189
OY 2181 GTGAGACCTTCGAATCTGTGACATATGACGATGACCTTAAGCTCAGATTTGCTCA 2240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2190 gtggtgcacattacgtcactgagttgacaaccacacatagatgtgtcacaattggggtta 2249
OY 2241 AAGGGTGTGTGAGATGCGATTTTGCATCGAAAAACCAAGGGCTTGTGCTATTGATG 2300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2250 ggggtctgtgggagagatgaagtgcttgcacacagaaacagttggtgcaactagatg 2309
OY 2301 GGGAGATGTTGGTTCAAGTATGATCAGACCAAAATGCTGTGTTCAATGCCATGCG 2360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2310 gggtagtctgaattgtatgtatg---agatlaaattgctgagagtgcaagttccctggtc 2366
OY 2361 CAATTGATTTCTTCATCG 2377
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2367 ctaagtctcaaaaatg 2383
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 6
AA220208
ID AA220208 standard; cDNA; 2690 BP.
XX
AC AA220208;
XX
DT 17-JAN-2000 (first entry)
XX
DE Sugarbeet raffinose synthase cDNA.
XX
KM Raffinose synthase; sugarbeet; transgenic plant; ss.
XX
OS Beta vulgaris.
XX
FH Key Location/Qualifiers
FT CDS 236..2587
FT /tag= a
FT /note= "this region (minus the stop codon) is
FT specifically claimed in Claim 5"
XX
XX EP953643-A2.
XX
XX 03-NOV-1999.
XX
XX 27-APR-1999; 99EP-0107430.
XX
XX 30-APR-1998; 98JP-0120550.
XX
XX 30-APR-1998; 98JP-0120551.
XX
XX 04-DEC-1998; 98JP-0345590.
XX
XX 10-DEC-1998; 98JP-0351246.
XX
XX (SUMO ) SUMITOMO CHEM CO LTD.
XX
XX Watanabe E, Oeda K;
XX
XX WPI; 1999-593144/51.

```

```

DR P-PSDB: AAY32073.
XX
XX New sense and antisense genes, useful for altering the level of
PT raffinose in food plants -
XX
XX Claim 5; Page 25-29; 55p; English.
XX
XX This is the nucleotide sequence of a sugarbeet cDNA clone coding
CC for raffinose synthase (see AAY32073), a protein which can bind a
CC D-galactosyl group through an alpha(1-6) bond to the hydroxy group
CC attached to the carbon atom at the 6-position of the D-glucose
CC residue in a sucrose molecule to form raffinose. The cDNA was
CC isolated from sugarbeet cv. haming leaf cDNA by PCR. Probes or
CC PCR primers generated from plant raffinose synthase genes (see
CC AAY20207-10) may be used to obtain other raffinose synthase genes
CC by labeled detection or amplification (claimed). These genes may
CC be used to control the levels of raffinose produced in plants.
CC Antisense genes can be used to knock out existing gene activity,
CC and sense genes to increase the level of gene activity. The
CC resulting transgenic plants may be used as a food source to alter
CC the growing conditions for gut enterobacteria, providing general
CC health advantages. The sense and antisense genes may also be used
CC in gene and phenotypic analysis of plants and for the selection of
CC plant clones with the desired characteristics with respect to
CC raffinose content.
XX
XX Sequence 2690 BP; 762 A; 507 C; 641 G; 778 T; 2 other:
SQ
Query Match 38.8%; Score 976.8; DB 20; Length 2690;
Best Local Similarity 67.5%; Pred. No. 3.5e-285;
Matches 1486; Conservative 1; Mismatches 683; Indels 30; Gaps 7;
OY 238 GGTTCGGTGGTTCCTTGTGATTCGACGCGTCCGACCTGATPAGCCGATGTTGT 297
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 403 ggggtctgtgggagagatgaagtgcttgcacacagaaacagttggtgcaactagatg 462
OY 298 TTCGATTTGGGAAGCTGAAGATATCGGTTTATGATTTTACAGTTTAAAGTTTGGG 357
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 463 atccgtggccagctccaaggaagatccctccatgagatctcctcaagttcaagatctg 522
OY 358 GACTACACACTGGTGTGCTGAATGCTGGGATCTGATGAGAGACTCAGATTTGTGAT 417
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 523 gactaccattgagctggttccaatgagcggaaccttgagcatgaccacaattctcat 582
OY 418 CCTTGAAGAAGTACA-----TTCTGTGTCGACCGTATGTTTCTTCTTCCATGCTTGA 471
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 583 ccttgataagtcagatgaagtggttggccgctcccatattgtgactccctccatgactga 642
OY 472 GGGACCGTTCCGAACTCGATTCAGCCTGG---GGATGATGACTTTGTCGATGTTGTT 528
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 643 agggccatttcgggcatccctccacgggtctctgtgactgactctgagatatactg 702
OY 529 CGAGATGTTTGGTGAAGTGTGATGATCGATTCGCAATATGTTGATCTTTCATGC 588
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 703 tgaagtggtgctcactaaagtgtcgtggaacctcgttcgggtctctctataatacg 762
OY 589 TGGTATGATCCGTTTGCATTGTTAAAGAGCGCATGAAGATGAGAGACCCATCTTGG 648
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 763 tgggcctgaccatttaagtttaataaagatacaatgaagagatcccaaccattagg 822
OY 649 AACTTTTCCTTGTGGAGAGAGATCCACAGGTATTCGTAATGTTGATTTGGTTGGT 708
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 823 gacttcaaaactctagatgacaaaactccctccagaaagatgagcaagttctgagctg 882
OY 709 CACGTGGACGCGTGTTCCTAAGCTTATCATCCAGAGCGCTGAATGAAGAGCGGTAGGCA 768
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 883 tacatgagatcatcttactcaaaagttagagcwcatalggtgttgggaaggttaaaag 942
OY 769 TCTGTGACAGCGGTTGTCTTCGCGGTTTACTCTAATCGACGATGTTGGCAATTCAT 828
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 943 actgctgaaaaaggggtcccaaccggtctcgtaactatgatagtgagtgagcaatctat 1002

```


QY	829	CGGACAGATTGGAAATCCCATCACA----	AAGAAGATGAACCAACACCTCCCGGGCA	885
Db	1003	ctgcatagcagatgataccgataatccacgaagaaggataaacccgactcttcgcgcga		1062
QY	886	GCAAAATGCCCTGGCCGCTCTTTTGAAATTC	CCAGAGAAATTACAAATTCCTGATCACTGCA	945
Db	1063	gcaaatgcatactgatactgatacgaagaaacttcaagtttaggactataaag		1122
QY	946	TCCCAAGGCCACCCGGCCCCGAG----	CCGGCCAGAAAGGGGATGGAAGCGCTTTATAGATGA	1002
Db	1123	cccaaatatattagggccatlgagatcatcccaataatggaatagggcccttcttaggga		1182
QY	1003	ACTCAAAAGGAGATTAAACATGCTGGAGCATGTTATATTTTGGCATCGCTTGTGTGATA		1062
Db	1183	cccttaaggaggagttccaacactgctgagcatgtgatactttaggactcttlaagggcta		1242
QY	1063	TTGGGGTGCCCTTCCGCCAGGTGCCTTGCTTGGCTTAGGCACTGTGATTCAGCCACT		1122
Db	1243	ctggggagggtlaaaggcccaatgctccagccctaccrryagggcccaagtagtaaaccccaaa		1302
QY	1123	GCATTACCAACGGGCTGCATGATCGATGTGAGATTTGGCCGGGATTAAGATGTTCTTCA		1182
Db	1303	gcttccccgggctctgagatgacaaatggaagatctagccttgagataaatactgtaataa		1362
QY	1183	TAAAGTCGGGCTGGTCCGCCGAGAGAGCTGAGGAGATGTACGAGACTTCATGCTCA		1242
Db	1363	tggatattggctcttcacagcctgataaaggcccaagaacttataggaaggttgcatactca		1422
QY	1243	TTTGGAAAAAGTTGGGATCGACGGCTTTAAATGATACCTTTATCCACCATTTGGAGATGT		1302
Db	1423	cttggaaaaacttggagatctgataggagtaagaattgatactcaaccttgcttggagatgact		1482
QY	1303	GTCGGAAGACTATAGGAGGAGAGATGTGATTTGGCAAGGCATTTACAAAGCAAGACAA		1362
Db	1483	ggcagagagactataggagggaagggttgaactagcaaaaactatactaataagcaataacaga		1542
QY	1363	ATCAATAATAAATTAACATTTTAAAGAAATTTGAGTCAATTGCAGTATGGAACATTTGTAACA		1422
Db	1543	atcagctgcgaagacattcaaaagcaacggttgatctgacgacagagacagtgacaaga		1602
QY	1423	CTTATGTTTCCCTTGGACAGCGAATCTTCTTGGTGTGTGGTGTGATTCCTTTGGTG		1482
Db	1603	ttctatgctccttggtaacagagacacattgctctgctgctggttgggttggactcttggcc		1662
QY	1483	CACGACCCCTCTGTGTGATTCACAAACGTAACGTTTGGCTCCAGAGATCTCAGATGTTCA		1542
Db	1663	aacgtacccgtctcgagagataataatgatacatattgctccaaaggcttctaatactgta		1722
QY	1543	TTTGGCAACAGACAGCTTGATGTGGAGAACTTCAATCCACCTGACTGGATATGTTCCA		1602
Db	1723	ctgagcccaaaatagacttaatggaatggaaaaacttatacaaccttgagtgagcaagttcca		1782
QY	1603	ATTCACCCACCCCTTGTGCCCCGCTTCCATGCTGCTCTCGAGGCATCTCTGGTGGCCGAT		1662
Db	1783	atctacaacaccttggctgtaatttcaatctgcatcctctggagacttcttgglygacaact		1842
QY	1663	CTATGTTAGTGAATCTGTGTGGGAAAGCATTAATCTTGAATCTTTCGAAAAAATATAGTCTTC		1722
Db	1843	ctatgattagatgatacttcttggcaagaataaacatcccttgctcaaaagctctgccttggcc		1902
QY	1723	TGATGATGATCATCTCTTGAGTGTGAGTACATATGACACTCCCGAATCCGATATGTTGTTTGA		1782
Db	1903	tgaatggttgcatactcttcgttggtagtaacatgacacttccctataagagattgcataattgt		1962
QY	1783	AGACCCCTTTGCATATATGAGAGAAACTATGCTTTAAGATTTTGGAAATTCACAAAGTTCACTGG		1842
Db	1963	agatcccttggacatgacaanaaactgtccaaatcttggaaacctccaacaagtaacaatgg		2022
QY	1843	AGTGAATGTGTCATTTCAACATGCCAAGAGAGATGAGTGTGCTGTGAGACACAGCCGCAACA		1902
Db	2023	agtccttggagcttctcaattctgcacaaggagagggtggagcgttgagcttcggaaaaaact		2082
QY	1903	ATGCTTTTACAAATATCTCAAAAGAGTGTGATCTCAAAATTAACCCCAAAAGATATGAAATG		1962

Db	2083	atgcttcacgagatctcaaaacctctccgcgaagaagatccaaagatglttgct	2142
Oy	1963	GCACAGTGGAGAAAACCCATATCTCTATTGAAGCGCTTAAACCTTTGGCTTTAOCCTTA	2022
Db	2143	ggagaaagagacacaagccattccccaatgaagagtgatgttctgcattactcac	2202
Oy	2023	TCAAGCCAAAACCTTATCTCTCTCCAAAGCCCTCTCAAGATCTTGACATAGCTCTTTGACC	2082
Db	2203	caaggaaaaaaactaactcctctcacaactatgtcaccatlgttaaatatacattgaccc	2262
Oy	2083	ATTGCAATTGGAGCTCATCTCTTTCACAGAGACCAACATCTCAACACTTCTCTACA	2142
Db	2263	cttcgatctcagacttatgttagtctctcccgatgataatcttacccttgaggatgcac	2322
Oy	2143	CTTTGGCCCCAATTGGCGTGGTGAACATGCTTAAACACTGATGAGCCATCTCTGCGA	2202
Db	2323	attgcaccccatagatagtagtaaacatgctcaacgcgcgaaggagcagctcaagctcttga	2382
Oy	2203	C-----TATGACGATGACTTAAGCTCAGTCCGAGATTTGGTGTCCAAAGGTTGGTGGACAT	2266
Db	2383	catcagtagagataatagataagaatagcttcaggttgtagtataaagagggccggagaat	2442
Oy	2267	GGCAGTATTTGGCATCGAAAAAACCAAGGGCTTCTCTATTGATGGGAGAGATTTGGGCT	2316
Db	2443	gattgttcaatcacagaagaagccaaagcgtgttagagtagtaatgagaagaagacatgagctt	2502
Oy	2317	CMAGTATGATTCAGGACCCAAATGTGTGGTGTTCAGATGCCATGSCCAATTGATTTCTTCA	2376
Db	2503	ttagatagaagaagcgc--atgattaaagttcaagtacatg-----aacataactc	2553
Oy	2377	GGGTGGCAATTTCCGTTATCGAGTACTTGTTTAAATTTTA	2416
Db	2554	agtgagtttaccacctgttagacttacttatgttagcttga	2593

RESULT 7

ID AAZ10002 standard; cDNA to mRNA; 2497 BP.

AC AAZ10002;

DT 26-OCT-1999 (first entry)

DE Nucleic acid encoding a raffinose synthase protein.

.....
KW Raffinose synthase; plant; sucrose; raffinose; ss.

Glycine max. OS

FH	Key	Location/Qualifiers
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

```
/*tag= a
```

XX:

XX

XX

XX :

PR 18-DEC-1996;

(SUMO) SUMITOMO CHEM CO LTD.
PA

DR WPI; 1999-51112/43.

XX 1

XX 5

XX

Db	2009	atcggtttgctgtatatttgltcaaggaccacaactaaagctatgsaagcatcaaga	2068
QY	2061	ATCTTGACATAGCTCTTTCAGCCATTTCGAATTCGAGCTCATCTGTTTCACCAAGTACCA	2120
Db	2069	aattggaagttcaactcttgagccaatttacttcttgagctatgacagtgctccagtgattg	2128
QY	2121	AACTCATCCAACTTCTCTACACTTTGCCCAATTGGCTGGGTGAACATGCTTAACTA	2180
Db	2129	tgctgtcaaaaaagtaattcaattcaattgtcccaattgattgagcaacgtcttaacactg	2188
QY	2181	GTGAGACCATCCAACTCTGTGAGACTAGAGAGAGACTAGTCAGTCGATGGTGTCGA	2240
Db	2189	gtgtgtcattcgcattgcattcgaatgagtttgacaaccaatagaatgtgtcaaaattggttla	2248
QY	2241	AAGGCTGTGATGATGCGAGATTTTGATATCGAAAAACAAGAGGCTGTGCTATTATG	2300
Db	2249	gggggtctgtggggagatgaagtgcttgcattcagagaaacacagtttggtgcaaacatg	2308
QY	2301	GGGAGAGTGTGGGTTCAAGTATGATGAGGACCAATGGTGTGTTCAAGTGCATGGC	2360
Db	2309	gggtgactgtcaaaattgtgattatg---agataaaatgtcgagatgtgcaagttccctg	2365
QY	2361	CAATTGATTTCTTCATCG	2377
Db	2366	ctagtgtctcaaatg	2382

RESULT 8
 AAH27438
 ID AAH27438 standard; DNA; 2343 BP.
 XX
 AC AAH27438;
 XX
 DT 17-AUG-2001 (first entry)
 XX
 DE Soybean polynucleotide; SEQ ID 2.
 XX
 KW Mutant; mutein; raffinose synthase; raffinose oligosaccharide reduction;
 KW plant; soybean; ds.
 XX
 OS Glycine max.
 XX
 PN JP2001078783-A.
 XX
 PD 27-MAR-2001.
 XX
 PF 03-JUL-2000; 2000JP-0200571.
 XX
 PR 09-JUL-1999; 99JP-0196036.
 XX
 PA (SUMO) SUMITOMO CHEM CO LTD.
 DR WPI: 2001-313373/33.
 DR P-PSDB; AAB98659.
 XX
 DT Novel mutant protein of raffinose synthase is useful for reducing the
 DT raffinose oligosaccharide content in a plant body -
 XX
 PS Disclosure: Page 20-23; 30pp; Japanese.
 XX
 CC The present invention relates to a mutant protein of raffinose synthase
 CC in which at least one aromatic amino acid present at the position of
 CC about 1-7 amino acids from the N-terminus is deleted or replaced. The
 CC mutant protein can be used for reducing the raffinose oligosaccharide
 CC content in a plant body. The present polynucleotide from soybean, was
 CC used in the present invention.
 XX
 SEQ Sequence 2343 BP; 597 A; 493 C; 648 G; 605 T; 0 other;

Query Match	38.4%;	Score 966.8;	DB 22;	Length 2343;
Best Local Similarity	66.0%;	Pred. No. 3.5e-282;		
Matches 1508; Conservative	0;	Mismatches 737;	Indels 39;	Gaps 6

QY	112	TGACATGTCGTCAACCGTTTGGCAATGCAAGCATGCAATTTCACTGTGAACGCTCATTCGTT	171
Db	57	taatttccttcttgcataaacccttagaagatacttaattcttcctgcgaagccacccctt	116
QY	172	TCGTGCGATGTTCTCGAAGCAATTGTTTCTCTCTCTCTCCGTMAC-----AC	219
Db	117	tctacaggaagttcccggaanaacataaagtaaccccttcaccatcaccatcgacgccaagtag	176
QY	220	TTTGATTTAGCAAGTCCCGGGTTTTCGGTTTGTTTGCTTTGGATTTCGACGCGTTCGAAC	279
Db	177	taagaacacaggaagaaagcaagcagtcgtatggtctcgtcgttcgaacgagcgagcc	236
QY	280	TGATAGCCGACATGTTGTTTTCGATTGGGAAGCTGAAGGATTTGCGTTTATGACTATTT	339
Db	237	cagaagccgaaacgcttgcttccttccttgaggagctcagagaaataatccatgacatatt	296
QY	340	CAGGTTTAAAGTTTGGTGGCATACACACTGGGTTGGTTCGAATAGTGGGGATCTTGATC	399
Db	297	ccggtttaaagttgtgtgagaccatacctacgttggtcgttagcaacggaacgtggaaca	356
QY	400	GGAGACATCAGATTGTGATTCCTTGAGAAGTCAG---ATTCTGCTCACCAGTATGTTTCT	456
Db	357	cgaagacaaagatctgactctcgcacaaaagacacacgctcgaagccctcttgctgatat	416
QY	457	TCTTCCGATCTGTTTAGAGGACCGCTTCCGAACCTCGATTACGCTGGGGATGATGACTTGT	516
Db	417	tctccgcgactccccaagccctcgtctcgagcctccctccgaacccggtttagatgatacgt	476
QY	517	CGATGTTTGTTGTCGAGTGTGTTCTCGAAAGTTGTTGATGATGATCGTTCCGACTATGTT	576
Db	477	ggaagcttgcataagagagcggtcgcgaacagtgctctcgttcagctcggagagctgctt	536
QY	577	GATCTTCATCCTGCTGTGATGATGATCCGTTTTCGACTGTTTAAAGGGGGATGCAATCGAC	636
Db	537	atacgtccacgttgcacatacgaaccggtatacagttgcttagaagaaactaaagtcgtag	596
QY	637	GACCCATTTGAGACTTTTGGCTTGTTGGAGAGAAAGATCCACACAGGATTCGTGCACAA	696
Db	597	gatgcatttgggagcttcaagcttctcgcagaggaagaaaccggtcagcagtgatcgaacaa	656
QY	697	ATTTCGTTGTCGACGTGGAGGCGGTTTTACTTAACGTTTATCCACAGGCGTAAATGA	756
Db	657	gttctgtgtggtacatcagagcaggttlaactlgaagtgatccctcaaggtgtgtg99a	716
QY	757	AGGCGTGAAGCATCTCGTCCACAGGCGGCTTCTCCCGGTTTAAAGCTAAACGACAGTGG	816
Db	717	aggaggtlgaagaggtgtgtgtlgaaggaaggtgtccctcagaagaaagtctcctaaccgaagcg	776
QY	817	TTGGGCAATTCACATCGACACGATTTCGGATCCCATCAC---CAAAGAAGAAATGAACCAAC	873
Db	777	gtggaagacacttgttcaagcagagagccccataacaggaacaaaggggtatagaagcgagac	836
QY	874	CGTGCSCGCGACAAATGCCCTGCGCTCTTTGAAATTCGAAGAAATTACAAATTCG	933
Db	837	ctccgcgaaggaagacaatgcatacgtlgtgtgaagttlgaaggaataatlaaagttcag	896
QY	934	TGACTACTCATATCCCAAGGCGACACGGCCGCCGACGGCGCAAGAAGGGATTAAGCGCTT	993
Db	897	acagatattgtatgtgaaagagttctg-----agaaggtatagtgtgctt	941
QY	994	TATTAATGTAACCTCAAGAGAGATTTAAGACTGTGAGCATGTTTATGTTTGGCATGCTTT	1053
Db	942	tgtttaggaacttgaagaaacagctttagagagcgtgagagcaagtgatgt---gcaagcgct	998
QY	1054	GTTGTGATATTTGGGGTGCGCTTGCGCCCGACAGTGCCTGGCTTGCTGAGGCACTGTGAT	1113
Db	999	tctgtgtatgtg99gtg99ggtcagaaccacaagttcccggtcagatgcacccaaggtcctaagttgt	1058
QY	1114	TGAGCCAGTGCCTTTCACAGGCGTCGACGATGAGAGATGAGAGATTGGCGGTGATTAACAT	1173
Db	1059	caactccgaagagtgctccaatgactcaaaattgacaatagaagatcttagcgtgtgataaagat	1118

Db 1059 cactcgaagctgtccaatgactaaattgacaatgaagatttagcggtgataagat 1118

QY 1174 TGTCTTCATAGAGTCGGGCTGGTCCCGGAGAGGCTGAGAGATGTACGAGACT 1233
 1119 cgtcagtaagagggatcggactcgtgcacacaccccgccaccccttcttgaaaggggctc 1178
 QY 1234 TCATGCTCATGGAAGAAAGTTGGATTCGAGGTGTTAGAGTATCCACTATT 1293
 1179 ccaaccgcgtcttgaaatcgcggtatcggtcgttaagcttcagactacacatcgtc 1238
 QY 1294 GGAGATGTTGTGTGAAGACTATGAGGAGAGATGTTGGCAAGGCAATTACAAAGC 1353
 1239 cgaagatgcataccgaggaalaacgctgcttgagctagccaaagctattacaagc 1298
 QY 1354 AATGACCAATCAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1413
 1299 gctcactgcctcgttgaaagagcattccaagcaatgggtcattcgaagccttgagagca 1358
 QY 1414 TTGTACGACTTATGTTCTTGGCAGGAGAGTATCTCTTGTGCTGTTGGTGTGATGA 1473
 1359 tctgaatgaactctctctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1418
 QY 1474 CTTTGTGTGACGAGACCCCTCTGTGATCCAAACGTTTGGCTTCCAGAGTGTCA 1533
 1419 tcttctgtgcagcagatccctcctcctcctcctcctcctcctcctcctcctcctcctc 1478
 QY 1534 CATGCTTATGTTGTCACAGCAGCTTGTGATGGGAGTATCATCCACCTGACTGAGGA 1593
 1479 catcgtgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1538
 QY 1594 TATGTTCAATTCACACCCCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1653
 1539 catgttcacagtcac 1598
 QY 1654 TGGCCGATCTATGTTAGTATGTTCTGTGGAAAGCATATTGATCTTGTGAAAAACT 1713
 1599 tggac 1658
 QY 1714 AGTGTCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1773
 1659 cgtcttcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1718
 QY 1774 TTTGTTGAAGACCTTGTGCTGATGAGAGAACTATGCTTGAATTTGATCTCAACA 1833
 1719 tctgtttgaagacccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1778
 QY 1834 GTTCACCTGAGATGTTGTGCTGATTCACCTGCAAGAGAGATGCTGTGTGAGACAGC 1893
 1779 atatacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1838
 QY 1894 CCGCAACCAATGCTTTCACATATACATAAAGAGTGCATCCAAACCTAACCAAGAGA 1953
 1839 gagaacaagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1898
 QY 1954 CATGAAATGCGACAGTGGAGAAACCTATCTATTTGAAGCGTTTAAACCTTTCGCT 2013
 1899 catgaaatgagcaatgagcaatgagcaatgagcaatgagcaatgagcaatgagcaatgagca 2013
 QY 2014 TTTACCTTATCAAGCAAAACTTATCTCTCCAGCCTCTTCAGAGATTTGACATAGC 2073
 1959 atattcttcaagaagcacaacaaagcctcagaagcagcagcaatgagaagcttc 2018
 QY 2074 TTTGACCAATTTGGAATTTGAGGCTGATGTTTCACGAGTGCACCAATTCATCCAAAC 2133
 2019 actgagacacattcattcttgagcattgagcagcagcagcagcagcagcagcagcagcagc 2078
 QY 2134 TTTCTTCACTTTGCCCAATTTGGGCTGTGTAACATGCTTAACTGATGAGAGATCA 2193
 2079 gtaaatcacaattctgacacattgagatgagcaatgagcaatgagcaatgagcaatgagca 2138
 QY 2194 ATCTGTGAGATTAAGCATATCAAGCTCAAGTGCAGATTTGGTCTCAAGAGGTGTGTA 2253
 2139 gtccaaatgagcttgcagcaac 2198
 QY 2254 GATGCGAGTATTTTCATGAGAAAAAACCAAGGCGTGTGATGATGGGAGATGTTGG 2313

Db 2199 gatgaagcgtctgcatcagagaacacagctagctcgaactcagatgagggctagctga 2258
 QY 2314 GTTCAAGTATGATCAGACCAATGTTGTGTTCAAGTGCATGCGCAATTTGATTTCTTC 2373
 2259 attgattatg---aggataaatgctgagagtgcaagctccctcgtgacctagctcctcaaa 2315
 QY 2374 ATGC 2377
 2316 atctg 2319
 Db 2316 atctg 2319
 RESULT 9
 AAV40800
 ID AAV40800 standard; cDNA to mRNA; 2746 BP.
 AC AAV40800;
 XX
 DT 23-SEP-1998 (first entry)
 XX
 DE Broad bean raffinose synthetase coding sequence.
 XX
 KW Raffinose synthetase; metabolism modification; food additive;
 XX gastrointestinal flora; broad bean; ss.
 OS Vicia faba.
 XX
 FH Key Location/Qualifiers
 FT CDS 101..2500
 FT /tag= a
 FT /product= raffinose synthetase
 XX
 PN EP849359-A2.
 PD 24-JUN-1998.
 XX
 PF 18-DEC-1997; 97EP-0122417.
 XX
 PR 18-DEC-1996; 96JP-0338673.
 XX
 PA (SDMO) SUMITOMO CHEM CO LTD.
 XX
 PI Oeda K, Wantanabe E;
 XX
 DR WPI; 1998-324670/29.
 DR P-PDB; AAM57886.
 XX
 PT New nucleic acid molecule encoding plant raffinose synthetase -
 PT capable of producing raffinose, used as food additives with
 PT beneficial effects on gastrointestinal flora
 XX
 PS Claim 1; Page 26-29; 44p; English.
 XX
 This sequence encodes the broad bean raffinose synthetase of the
 CC invention. The raffinose synthetase is capable of producing raffinose by
 CC combining a D-galactosyl group through an alpha (1-6) bond with a
 CC hydroxyl group attached to the carbon atom at position 6 of a D-glucose
 CC residue in a sucrose molecule. The DNA can be used to modify metabolism
 CC of a host organism by introducing into the host organism or its cell so
 CC that the content of the raffinose family oligosaccharides in the host
 CC organism or cell is changed. Raffinose family oligosaccharides are useful as
 CC food additives with beneficial effects on the gastrointestinal flora.
 XX
 Sequence 2746 BP; 780 A; 552 C; 621 G; 793 T; 0 other;
 SQ
 Query Match 38.0%; Score 956.6; DB 19; Length 2746;
 Best Local Similarity 66.6%; Pred No. 4.8e-279;
 Matches 1421; Conservative 0; Mismatches 694; Indels 18; Gaps 3;
 QY 248 GGTGCTTGTGATTCGACGCGTGGAGAACCTGATAGCCGACGATGTTGATGATGGG 307
 1111 11 111111 11 1111 1111 11 1111 11 1111 11 1111 11
 Db 359 ggtgtgtcgttgattcactccacgacacacacacacacacacacacacacacacacacac 418

[illegible]

QY	655	GAGGAGAAAGACTCCACCAGATATCGTGGACAAATTGGTGGTGCACGTGGAGACCGCTTT	724
Db	779	gaaagaaaaaacagacacclatlatatagaacaaattcgtgttgcacgttggaaatgccttc	838
QY	725	TTACCAACGGTTCATCCACAGCGCGCTTAATGAAAGGGGTGAGGCATCTGCTGCACGGCGGT	784
Db	839	tacttgaagttctccaaaaagtgatctggaaagtgtaaaagtctctccaaagatggtgtc	898
QY	785	TTGTCCCTCCCGGTTTAAGTCTTAATCGACGATGTTGGCAATTCATCGGACAGATTCCGAT	844
Db	899	tgtctcccggttttcgtcataatccgacggtgtgcatactcatgtcatatgaaatgac	958
QY	845	CCCATCCACCAAAAGAGATGAACCAACCGTCGCGCGGACCAATGCGCTGCCGCTTT	904
Db	959	gatbaagatgatctaaagaaatgaaacccaaacccacgcggygaaacaaatgcacatgcagactt	1018
QY	905	TTTGAATTTCCAAAGAGATTTCAAAATTCGGATACGTACTGTCAATCCCAAGGCCACCGGCCCC	964
Db	1019	gtaaaaatcgaagaagaaatctctaagttttagagaataatgaaatccgcga-----a	1066
QY	965	CGAGCCGCGCCGAAAGGGGATGAAGGGCGTTTATATGATGAACCTCAAAAGAGATTAAAGCT	1024
Db	1067	aatgagaggaaagaagttcgtgggtcttctgtgaggaattctgaagaaagatcttggagtc	1126
QY	1025	GTGAGACATGTTTATGTTTGGCATGCTTTGTGTGATATTTGGGGTGGCCTTCCGCCGAG	1084
Db	1127	gtggaagtgttctatgtttctgtgcacgcgcttctgggtatcttgggacggggtcttgccttga	1186
QY	1085	GTGGCTGGCTTGGCCTGAGGGGACGAGTGATTCACCGACGTCCTTTCACAGGGCTGCAAGATG	1144
Db	1187	gtgcattggaatgcgcgaagaagctgaaggttctgtctcgaagagtgctctaaagggtcttgaagatc	1246
QY	1145	ACGATAGAGATTTTGGCGGTGAGATTAAGATTGTTCTTATAAAGGTGGCGCTGCTCCCGCG	1204
Db	1247	acgatgagagatcttgcggtggtgataagattgttgaabaaggtggtgggtctagtcgcgcga	1306
QY	1205	GAGAAAGCTGAGAGATGTACAGAAAGCACTTCATGCTTCATTTGGAAAAAGTTGGATGCAC	1264
Db	1307	gatttgcacatgaaatgltcttgaatggcttcaactcctcatcttggatcggtcggyaaattgac	1366
QY	1265	GCTGTTAAGATTTGACGTTTATCCACTATTGGGAATGTTGTGTGAAGACTAATGAGAGAGA	1324
Db	1367	gggtctaaagtttgatctcatccatctgcgtgagttacatacgaagaaatgaatgtagaagca	1428
QY	1325	GTGGTGGTGGCAAAAGCATATTACAAAGCAATGACCAAAATCAATAATTAACATTTTAAA	1384
Db	1427	gttgcagctagaagaagcttatatacaaaagcctaactcatcgtgtagaagaaacaaatccaa	1486
QY	1385	GGAAATGAGATCATTTGCAAGATATGGAACATTTGTACGACTTCATGTTCTCTTGGCAGGAA	1444
Db	1487	ggcaatggttgaatttgcctagcatgtagagcatltgaacagacttctctccctgcgcga	1546
QY	1445	GCTATCTCTTGGTCTGTGTTGGTGTGATGACTTTTGGTGCAGGACCCCTCTGTGTATGCA	1504
Db	1547	gocatatccctcgcgcgcgtccgagatgattcttggctgcgtcatccatctgtagatcca	1606
QY	1505	AACGGTACGTTTGGGTCCAAAGATGTCAATGTTTCATTTGGCCAAACAGACGTTTGGG	1564
Db	1607	aatgttaataatgtgcctccaaagttgtccaaatgtaacatltgtagcctacaacaaacggttla	1666
QY	1565	ATGGGGAATTCATCCACCGCTGACTGGGATATTTCCAAATCCACCCACCCCTGTGCGCGC	1624
Db	1667	atgggaataattcatctacgtccgaatgtggaaatgtttccagtcacctcaatccctgtgciga	1726
QY	1625	TTTCATGCTGCTTCCAGACCATCTCTGTGTGGCCCAATCTAATGTTAAGTATCTGTGGGA	1684
Db	1727	tttcatgcgcctcaacagacatccgcgcgaacaaatlaatgtagtgaatgtgtgtc	1786
QY	1685	AAGCATACCTTTGATCTCTTGAAAAAACTGTGTCCTTCCTGATGAGATGCAATCTCTGGAAGT	1744
Db	1787	aatccaatctcaagttgcctcaaatctctgttcttgcgcgaatggtctctatctctgcgtgtc	1846
QY	1745	GAGTACTTGTGCACCTCCGACTCGGATGTTGTTCTTTGAAGACCCCTTGGCATATATGAGAA	1804

Db 835 agtlacagagagaattacaagtltaagagcta-----t 867
Qy 968 GCCCGCCAAAGAGGATGAAAGCGTTATAGATGAACCTAAAGAGAGTTTAAGACTGTG 1027
Db 868 aaggaaggaaggaaggggttgaagagagattgaagagagatttggtgtgtgtg 927
Qy 1028 GAGCATGTTTATGTTTGGGATGCTTGTGTGATATTTGGGGGCGGCTTGGCCGAGAGTG 1087
Db 928 gaggtaegttgaagttgagcaagcgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 987
Qy 1088 CCGTGGCTTGGCTGAGGACAGGTGTGATTCAGCCAGTGCCTTTACACGAGGCTGAGATGACG 1147
Db 988 gagggaagagcgagagcggt 1047
Qy 1148 ATGGAGAGATTTGGCGGTGATGAAGATTTGTTCTTATTAAGTGGGCTGTGCTCCGCGGAG 1207
Db 1048 atggaagagatcgt 1107
Qy 1208 AAGGCTGAGAGAGATGATGACAGGACTTGTGATTCATTTGGAATAAATGGATGACAGGT 1267
Db 1108 cgt 1167
Qy 1268 GTTAAAGATTGACGTTATCCACCTATTGAGATGTTGTGTGAAGACTATGAGAGGAGAGTG 1327
Db 1168 gtcaaaagt 1227
Qy 1328 GATTGGCAAGGCAATATTAACAAGCAATGACCAATCAATAATTAACATTTTAAGGA 1387
Db 1228 gataatgt 1287
Qy 1388 AATGAGATTTGCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1447
Db 1288 aagcggt 1347
Qy 1448 ATCTCTTGT 1507
Db 1348 atatccctgt 1407
Qy 1508 GGTACGTTTGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1567
Db 1408 ggtacatctgt 1467
Qy 1568 GGGAGATTTGCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1627
Db 1468 ggtacatctgt 1527
Qy 1628 CATGCTGCTGT 1687
Db 1528 catgt 1587
Qy 1688 CATAACTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1747
Db 1588 cacaatctgt 1647
Qy 1748 TACTATGACATCCGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1807
Db 1648 caetatagt 1707
Qy 1808 ATGCTTAAAGATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1867
Db 1708 atgtctaaagatgt 1767
Qy 1868 GAGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1927
Db 1768 gagggaaggt 1827
Qy 1928 GTGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1987
Db 1828 gtctac 1887
Qy 1988 ATTGAAGCGCTTAAACCTTTGGCTTAACTTATCAAGCCAAACCTTATCTCTCTCC 2047
Db 1888 attgaaaggggt 1947

Qy 2048 AAGCCCTCTCAAGA---TCCTGACATAGCTCTTGACCCATTCGAAATTCGAGCTCATCT 2104
Db 1948 gacacatctgt 2007
Qy 2105 GTTTCACACAGTACCAAACTCTCAAACTCTCTACACTTTGGCCCAATTGGGCTGTG 2164
Db 2008 gtttccctgt 2067
Qy 2165 AACATGCTTAAACACTATGTGAGACCATTCATCTGTGACTATGACGATGACCTAAGCTCA 2224
Db 2068 aatagtctaaacacgt 2127
Qy 2225 GTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2284
Db 2128 gtltgaagttgt 2187
Qy 2285 GCTTGTGCTATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2344
Db 2188 acgtgtgaagattgt 2244
Qy 2345 GTTCAAGTGCATGCGCAATTGATCTTCTATCG 2377
Db 2245 atccaagtacatgt 2277

RESULT 12
AAV40802
ID AAV40802 standard; cDNA to mRNA; 1762 BP.
XX
AC AAV40802;
XX
DT 23-SEP-1998 (first entry)
XX
DE Japanese artichoke raffinose synthetase coding sequence.
XX
KM Raffinose synthetase; metabolism modification; food additive;
KW gastrointestinal flora; Japanese artichoke; ss.
XX
OS Stachys sieboldii.
XX
FH Key Location/Qualifiers
FT 2..1762
FT CDS /*tag= a
FT /product= raffinose synthetase
FT /note= "no stop codon given"
XX
PN EP849359-A2.
XX
PD 24-JUN-1998.
XX
PF 18-DEC-1997; 97BP-0122417.
XX
PR 18-DEC-1996; 96JP-0338673.
XX
PA (SUMO) SUMITOMO CHEM CO LTD.
XX
PI Oeda K, Wantanabe E;
XX
PI WPI; 1998-324670/29.
DR P-PSDB: AAM57888.
XX
PT New nucleic acid molecule encoding plant raffinose synthetase -
XX capable of producing raffinose, used as food additives with
XX beneficial effects on gastrointestinal flora
XX
PS Claim 1; Page 36-38; 44pp; English.
XX
CC This sequence encodes the Japanese artichoke raffinose synthetase of the
CC invention. The raffinose synthetase is capable of producing raffinose by
CC combining a D-galactosyl group through an alpha (1-6) bond with a
CC hydroxyl group attached to the carbon atom at position 6 of a D-glucose
CC residue in a sucrose molecule. The DNA can be used to modify metabolism

Query Match	35.2%	Score 866.8	DB 19	Length 1762
Best Local Similarity	70.0%	Pred. No. 5.2e-258		
Matches 1223	Conservative	0	Mismatches 517	Indels 6
Gaps				
QY	379	AAATGATGGGAGCTTGAATCGAGACTCAGATTGTGATCTCTGGAAGTCAGATTCTGG	438	
Db	4	aaacgggtcgtactcttgaggggaaactcaaatagtcgtcgtcgaacgtccga---cga	60	
QY	439	TCGACGATAGTTTTCTCTTCTTCGATCGTTGAGGACCGTTCCGAACTCGATTACACC	498	
Db	61	cagcgccctcaatcgtctcgtcttcgcgtccacaggggcaggttcggtctcccttcacgc	120	
QY	499	TGGGGATGATGATCTTGTGCAATGTTGTGTCGAGATGAGTTGTGTGCAAAAGTTGTGAGC	558	
Db	121	cgggtggaagatcattacgcataattcgtgcgaagcgggtcaaccaagyltaacagatc	180	
QY	559	ATCTCTCCGAAGTATGTTGTATCTTCATGCTGGTGAATGATCCGTTTGACATTGTTAAAGA	618	
Db	181	ctcgtccgcgtcttcgctctacatacgaacgcgggtgtagcccttattaccccttgtaga	240	
QY	619	GGCGATAAAGTCGTGAGACCCATCTTGGAACTTTTCCTGTTGGAGAAACAGTCC	678	
Db	241	cgcgcgtgaagltgcgcgcacacccctcggagacgttcagtcgctggaagaaacatcc	300	
QY	679	ACCAAGATCGTGAGCAAAATTCGTTGTTGTCACAGTGGAGCGCGTTTAACTTAACGTTCA	738	
Db	301	gcgcgggatatcgtcgaacaattcgggtgagcagtgagtgagtgcttactcaacgtcca	360	
QY	739	TCACAGAGCGCTAATAGAGCGGTGAGGCACTCTCGTCACAGCGCGGTTGTCTCCCGGTTT	798	
Db	361	gcccaacgcgtttagaaggcgctgcaggggcgtgtgttagcgcgcgcgcgcgcgcgcgc	420	
QY	799	AGTCTTATCGACGATGTTGGCAATCCATCGGACAGATTGGAGTCCATCAACCAAGA	858	
Db	421	gggttcatcgaacgagcggtgcagctccattcgcagaaacagcgctcaacacgga	480	
QY	859	AGGATGACCAACCAACCTCCCGCGGACGCAAAATCCCTGGCGTCTTTGAAATTCACAGA	918	
Db	481	ggggatgggaggaacccctcgcgcggagagcaaatgccttcgaggttagtaaglttaga	540	
QY	919	GAATTACAAATTCCTGATACGTAACGTAATCCCAAGGCCACCGCCCGGACGGCCAGAA	978	
Db	541	gaattacaagttcaggagtagtaccgagcccgaaataaactcgggc---gggcccgaatac	597	
QY	979	GGGGAATGAAGCGCTTATAGATGAATCAAGAGAGAGTTTAATACATGTGAGACATGTTA	1038	
Db	598	ggggatcggggccttattcgtgcacatgaagagcaaatltaagagtggtgactaagtgta	657	
QY	1039	TGTTGGCATGCTTGTGTTGATATTGGGTTGGCTTCGCCCGCAGGTGCTTGCTTGCC	1099	
Db	658	cgtggtgagatcgtgtgtgtgtattatgggcgcgggtcctcagcccaatgttcgcggcctgc	717	
QY	1099	TGAGGCACGTTGATTCAGCCAGTGCCTTTACACAGGCTCTCAATACATGAGAGGATTT	1159	
Db	718	cgaagctcaagctatctagaccccaactgactcttcgggtcttaagacacacatgagaagat	777	
QY	1159	GGCGGTGATTAAGATTGTGTTCTTAATAGATGTCGGGCTGATCCCGCCGAGACGCTCAGGA	1219	
Db	778	ggcgtgtgataagatctcaacaatggcgtgggtcgtgtccacccggagttgttgtaaca	837	
QY	1219	GATGTACCAAGAGACTTCATGCTCATTTTGGAAAAAGTTGGATGCAGCGTGTTAAGATTGA	1279	
Db	838	aatgataigaagattacattcacatccgatatcgtgtgggattgtagtgcacaagtga	897	
QY	1279	CGTTATCCACCTATTGGAGATGTTGTGTGAAGACTATGAGAGGAGATGTGATTGGCAAA	1339	

Db	898	cgctacaccattctgttgaaatglttgtgaaagacatgttggggagtgagccttaagccaa	957
Oy	1339	ggcattatattacaaagcaatgaccaaatcaatataatataacatttttaagcaaatgacgcat	1398
Db	958	ggtctattacaaggccttataagctcaagctcaagtltaacaacacttcaacggtcaacggtcat	1017
Oy	1399	tgcacgatgatggaacattgttgaacgacttgcattgttcccttggacacggaaagctatctctttgg	1458
Db	1018	cgctggtccttgagcaactgcaatgacttcaatgttcttcggaacacggagccattaccttgg	1077
Oy	1459	tccggttggtggaatgacatttttgggacgagaccctctggtgcatccaaacgggtacgttttg	1518
Db	1078	tcggtctcggggaaatgttlttgtgtcacatgccatcttgtagagatcccaatbgtcacgttctg	1137
Oy	1519	gcttccaaagatgtcacatggtttcttattgttgccaaacgaacacttgttggaatgggaaatttcat	1578
Db	1138	gttgcaagaggtgtcacatgltgtaactgtgtaactgcgcctacaacagcatatgatatglaattcat	1197
Oy	1579	ccacacctgacatgsggattgtgtccaaatccacccttggccgcttccaaatggccctc	1638
Db	1198	ccacacctgatttggaatgttlltaactgcaatcccttgcgttgtaatttccacgcgtccctc	1257
Oy	1639	tccgagccatctmccggngcccgccgcatgtatgttattgtagtcttgttgggaaagacataacttga	1698
Db	1258	acgaagccatctccggcggggcccatttaagtcagtgactgcgtgcggaagaagcaacttcga	1317
Oy	1699	tctttctgaaaaaaactagtcgttccctgatgatgcattccttgcgaactgactatgcat	1758
Db	1318	gtctccttaagagctcgtctcttccegaatgctccatccctcgtltgtatctactaagcgtc	1377
Oy	1759	ccgcacatccggattgtttgtttggaagaccctttgcatatggagaacgataatgcttaagat	1818
Db	1378	tcgcacttcgcatgtccctctcttgaagatccacttcaactcaatggaagactatagtccaat	1437
Oy	1819	tttgaaatctcaacaaagtgtcactggaatgtattggtgcatttcaactgccaaagagagatg	1878
Db	1438	tttgaattataacaagatccaagtcacggagttgtgcgaacttccaactgccaaagtgtgcggtg	1497
Oy	1879	gtgtcgtgagacacacgccgcaaaccaatgctttttacaaatfactaaaaacagatgacatccaa	1938
Db	1498	gagccgggaagtgtgtcgcacaacaaatgtgcgtgcgagatattccacgcgtctcctctag	1557
Oy	1939	aactraccccaaaagacatgatatgaatggcacaactgagaaaaacccatctctatttgaagcgt	1998
Db	1558	cgcgtgtcgcggtgacatgtgagtggaagcaagaaacgagatccgatcgcagctgcgaagcgt	1617
Oy	1999	ttaaaaccttttcggttttacccttcatcaaggccaaaaaacttatctctctccaaagcctctca	2058
Db	1618	caaaaacatctcggtgttactcatctatccacggaagaagaactcgtctcttctlaagccatcaga	1677
Oy	2059	agatcttgacatgactcttggaccatcttgcgaatttcgagcttcatcacactggttgcacgctgac	2118
Db	1678	caaaatcgacatcagctgtgagccttcgatttggatcgtgataaacggtttctccagtcaa	1737
Oy	2119	caaaact 2124	
Db	1738	aactct 1743	
RESULT 13			
AAZ10003			
ID	AAZ10003 standard; cDNA to mRNA, 1762 BP.		
XX	AAZ10003;		
XX	26-OCT-1999 (first entry)		
DT	Nucleic acid encoding a raffinose synthase protein.		
XX	Raffinose synthase; plant; sucrose; raffinose; ss.		
KW	Stachys sieboldii.		
OS			

XX	Key	Location/Qualifiers
FT	CDS	2..1762
FT		/*tag= "raffinose synthase"
FT		/product= "no termination codon given"
PN	JPI1215984-A.	
XX	10-AUG-1999.	
PD	12-DEC-1997:	97JP-0342899.
XX	28-NOV-1997:	97JP-0329006.
PR	18-DEC-1996:	96JP-0338673.
XX	(SUMO) SUMITOMO CHEM CO LTD.	
XX	WPI: 1999-51112/43.	
DR	P-PSDB: AAY30144.	
XX	New raffinose synthase gene - is prepared from a plant material	
PS	Claim 13: Page 31-34: 40pp: Japanese.	
XX	The present sequence encodes a raffinose synthase protein. The	
CC	sequence is isolated from plant material. The protein forms raffinose	
CC	by complexing alpha(1 to 6)- D-galactosyl hydroxyl group of the 6C	
CC	of D-glucose residue in sucrose molecules.	
XX	Sequence 1762 BP; 395 A; 451 C; 499 G; 417 T; 0 other:	
XX	Query Match	35.2%; Score 886.8; DB 20; Length 1762;
XX	Best Local Similarity	70.0%; Pred. No. 5,2e-258;
XX	Matches 1223; Conservative	0; Mismatches 517; Indels 6; Gaps 2;
OY	379	AAATGGGAGGATCTGAATCGGAGACTGATGTCCTGAGAAATGATCTGCG 438
DB	4	aaacgggctcgactctcgcggaacacaaatagctgctgcacacgcga---cga 60
OY	439	TCGACCGATATGTTTCCTTCCTTCGATCGTTGAGGAGCCGTTCCGAACCTGATTCAGCC 498
DB	61	cagcgccatacactgctgctgctcgcctacacgagggcaggttcgcctccctcagcc 120
OY	499	TGGGAGATGACTTGTGCGATGTTGTGTCGAGAGTGTTCTCGAAAGTGTGATGC 558
DB	121	cggctgagatgacttactcagatattgctgcgaagcgggttcaacaaagctcaacgagtc 180
OY	559	ATGGTCCGAAAGTATGTTATCTTATGTCGATGTCGATCGCTTTCGCAATGCTTAAGA 618
DB	181	ctgcttcgcgctcgcctcctacacacgcgcgctgctgaccccttacccttgtagaaga 240
OY	619	GCGGATGAAGATCGTGAAGACCATCTTGTGATCTTTCGCTTGGAGAGAGAACTCC 678
DB	241	cgcgcggaagggcgccgcacacacccctcgagctcagctcgtgaggaagaaattcc 300
OY	679	ACCAAGTATCTGACAAATTCGTTGCTGACGTCGAGGAGCGCTTTTAAACGGTTCA 738
DB	301	gcccggagatctgcacaaatctcggtgctgacgctgcttaccctcaacgtccca 360
OY	739	TCACAGGCGGTATAGAAAGCGTGAAGGATCTGTCGTCGAGGCGGTTGTCCTCCGGTTT 798
DB	361	gcccaacggctctcagagggcgctgcagggcgctgctgacgagcgagatgcgcgggct 420
OY	799	ACTCTAATGACAGATGTTGGCAATTCATCGGACAGATTCGATCCATCAACAAAGA 858
DB	421	gggtctgctgacgacgagggcgagctcatttgcaagaaacgagcgcgtccacacga 480
OY	859	AGCAATGACCAACCGTCCCGCGAGCAATGCGCTGCGCTTTTGAATTTCCAGA 918
DB	481	ggggatgaggagaaactccgcgcgagagcaaatgcctcaggttgatcaagtttgagga 540

OY	919	GAATTACAATTCGGTACTAGTCATCCAAAGCCACCGCCCGACCGCCAGAA 978
DB	541	gaattacaagttcagggagatcagagcccgaaataaacttgccc---ggcccgatcac 597
OY	979	GCGGATGAAGCGGCTTTATAGATGAATCAAGAGAGATTTAAAGTGTGAGCATGTTTA 1038
DB	598	ggggatgaggcgcttactcgtgacacagaaatctcaagagtgctgacagtgta 657
OY	1039	TGTTGGCATGCTTGTGTCGATATTTGGGGTGGCTTCCCGCCAGAGTCCCTGCTTGGC 1098
DB	658	cgctgctgacgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 717
OY	1099	TGAGCGACGCTGATGATCAGCAATGCTTTCACGAGCGCTGACATGATGAGATGATTT 1158
DB	718	cgaagctaaagctcattgagcccaactgactcctggtgcttaagaccacatggaagatt 777
OY	1159	GCGGATGATTAAGATGTTCTTCATTAAGGTGGGCTGCTCCCGCCGAGAAAGCTGAGA 1218
DB	778	ggctgctgacagatgctgacacacatgctgctgctgctgctgctgctgctgctgctgac 837
OY	1219	GATGTACGAAGACTTCATGCTCATTTGGAAAAAGTTGGGATCGACGCTTAAAGTTGA 1278
DB	838	aatgataagagatctacatcactcgaatctgctgagagatgactgagatcgaagtgta 897
OY	1279	CGTTATCCACCTATTTGAGATGTTGTGTAAGACTATGAGAGGAGATGATTTGGCAAA 1338
DB	898	cgtaaccatctgttgagaaatctgtgtaagacatctgctgagagagtgacttaccga 957
OY	1339	GCGATTTACAAAGCAATGACCAATCAATTAATTAACATTTTAAAGAAATGATGTCAT 1398
DB	958	ggcttatacaagcgcttatacaagctcagtaacacacactcaacgagaaagcgctcat 1017
OY	1399	TGCAATATGCAACATTTGTAACACTTCAATGCTTCCTTGGCAGCGAATCTCTCTTG 1458
DB	1018	cgtctgcccggagacactgaaatgacttctcctcggaaacgagacacttacccttgag 1077
OY	1459	TGCTGTGTTGATGACTTTTGTGTCAGGACCCCTCTGGGATCAAAAGCTGACTTTTG 1518
DB	1078	tgtgtcggggagatcttctgtgacatgactcactgagatcccaatgagcagcttcg 1137
OY	1519	GCTCAAGATGTCACATGCTTATTTGTGCAAGACGACTGTTGATGGGAGAACTTCAT 1578
DB	1138	gttgcaaggggtgcacatgctgacacgcgcctacacacagacataatgagtgtaattcat 1197
OY	1579	CCACCTGACTGGGATATGTCATTCACACCCCTTGTGCGCTTTCATCTCTCTC 1638
DB	1198	ccacccctgactgagacatgcttcaatcgaactccctcgcctgactgactcgaacgcgtcctc 1257
OY	1639	TGAGCCATCTCTGTCGCGCCGATCTATGTTAGTATGTCGTCGTCGTCGTCGTCGTCGTCG 1698
DB	1258	acgagccatctccggcgagggcccaattacgtcagtgactcgctgcggaagacacacttga 1317
OY	1699	TCTTGTGAAAAAATAGTGTCTCTGATGATGATGATCTTTCGAAAGTGAATCTTCACT 1758
DB	1318	gtcctctgaagagcctgcttctccgactgctcactcctcgtctgacttaacagcgctc 1377
OY	1759	CCGCACTCGGATGTTGTTGTTGAAGACCTTTCGTAATGAGAAATCTATGTTAAGAT 1818
DB	1378	tccgactcggatctgctcctcttgaagatcccaactcacaagaaactactgctcaaat 1437
OY	1819	TTCGAATCTTCACAAATTCCTGAGTGTGTCGATTCATCTGCAAGAGAGAGAGATG 1878
DB	1438	ttagaattataaacaagttccacggagatctgcgaaacttcaactcgaaggttgcgagg 1497
OY	1879	GTCGTGACACAGCCGCAACCAATGCTTTTTCACAAATCTCAAAACGATGATCCAA 1938
DB	1498	gagccggagagtgctgcgaacacatgctgcgaggtatctccacgagcgtctccctcag 1557
OY	1939	AATTAACCCAAAGATATGAATGACCACTGAGAAACCTTATCTTATTAAGCGCT 1998
DB	1558	cgctgctcggatgactgactgaggaagaaagatccgactgacgctgacggcg 1617
OY	1999	TAAACCTTTGCGCTTATCTATCAAGCAAAACCTTATCTCTCCAAAGCCTCTCA 2058


```

Db      949  gagttccatcgtcgtcccaacgctgcacatccggtgagcccatccatccagcgtttgtgt 1008
QY      1682  GGAAGACATATCTTATATCTTCTGAAAAAACTGTGCTTCTCATGATGATGATCTTCCA 1741
Db      1009  ggcacacacagatctcgtcctccttgagagagctcgtcttctgacggttcgactttgaag 1068
QY      1742  AGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1801
Db      1069  tctgagtaacatcgtcctcccaacgtcgtacggtcctcttgagagacccctcctcgtatgac 1128
QY      1802  GAAATATGCTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1861
Db      1129  aaaaacatgctcaagatcttggaacttgaaacagtaacatcgtgaatcgtcgaatccgaatccac 1188
QY      1862  TGCCTAAGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1921
Db      1189  tctcaagagagagatctggtcagagaaactcgacgacccaatcgtcctcccaatcgtt 1248
QY      1922  AAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1981
Db      1249  aacacgttcaacgcgcgaacaaacccaatcgaatcgaatcgaatcgaatcgaatcgaatcgaat 1308
QY      1982  ATCTATATGAAAGCGCTTAAACCTTTGCGCTTACCTTATCAAGCAAAACCTTATC 2041
Db      1309  atctccatgtaaaacgttgagaggttctgtcttctgtcctcaatccaaagaggttgtg 1368
QY      2042  CTCTCCAAAGCCCTCTCAAGATCTTGAATGATGATGATGATGATGATGATGATGATGATGATGAT 2101
Db      1369  tctgcgaggaacaaacgaatcgtcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatc 1428
QY      2102  ACCTGTTTACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2161
Db      1429  acgtctcaccacgtctgcacatctgagagagatcgttcaatgctcgaatcgaatcgaatcgaat 1488
QY      2162  GTGAACATGCTTAAACCTTAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2221
Db      1489  gtaacatctgttaacatcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 1542
QY      2222  TCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2281
Db      1543  tccgttgaatcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 1602
QY      2282  AGGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2341
Db      1603  gtcagctgcaagatctgagtgtaagatctgagtgtaagatctgagtgtaagatctgagtgtaag 1659
QY      2342  GTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2401
Db      1660  atggttcaagatcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 1710
QY      2402  TGTGTTTAAATTTTAAATTTTATGT 2423
Db      1711  tctgttcaagatcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 1732

RESULT  15
AAZ20207 standard; cDNA; 928 BP.
AAZ20207:
AAZ20207:
XX      17-JAN-2000 (first entry)
XX      Soybean raffinose synthase cDNA.
XX      Raffinose synthase; soybean; transgenic plant; ss.
XX      Glycine max.
XX      Key Location/Qualifiers
XX      FT 2..799
XX      CDS /*tag a

```

```

XX      EP953643-AZ.
XX      PD 03-NOV-1999.
XX      PF 27-APR-1999; 99EP-0107430.
XX      PR 30-APR-1998; 98JP-0120550.
XX      PR 30-APR-1998; 98JP-0120551.
XX      PR 04-DEC-1998; 98JP-0345590.
XX      PR 10-DEC-1998; 98JP-0351246.
XX      PA (SDMO) SUMITOMO CHEM CO LTD.
XX      PI Watanabe E, Oeda K;
XX      DR WPI; 1999-593144/51.
XX      DR P-PSDB; AAY32072.
XX      PT New sense and antisense genes, useful for altering the level of
XX      raffinose in food plants -
XX      PS Claim 3; Page 20-22; 55pp; English.
XX      CC This is the nucleotide sequence of a soybean cDNA clone coding
XX      CC for raffinose synthase (see AAY32072), a protein which can bind a
XX      CC D-galactosyl group through an alpha(1-6) bond to the hydroxy group
XX      CC attached to the carbon atom at the 6-position of the D-glucose
XX      CC residue in a sucrose molecule to form raffinose. The cDNA was
XX      CC isolated from soybean cv. Williams 82 immature seed cDNA by PCR.
XX      CC Probes or primers generated from plant raffinose synthase genes
XX      CC (see AAZ20207-10) may be used to obtain other raffinose synthase genes
XX      CC by labeled detection or amplification (claimed). These genes may
XX      CC be used to control the levels of raffinose produced in plants.
XX      CC Antisense genes can be used to knock out existing gene activity,
XX      CC and sense genes to increase the level of gene activity. The
XX      CC resulting transgenic plants may be used as a food source to alter
XX      CC the growing conditions for gut enterobacteria, providing general
XX      CC health advantages. The sense and antisense genes may also be used
XX      CC in gene and phenotypic analysis of plants and for the selection of
XX      CC plant clones with the desired characteristics with respect to
XX      CC raffinose content.
XX      SQ Sequence 928 BP; 266 A; 171 C; 222 G; 269 T; 0 other;

Query Match 12.1%; Score 303.8; DB 20; Length 928;
Best local Similarity 63.0%; Pred. No. 1.8e-81;
Matches 486; Conservative 0; Mismatches 282; Indels 3; Gaps 1;

QY      1600  CCAATCACCACCCCTTGTGCGCTTCCATGCTGCTCGAGACCATCTCTGTGCGCC 1659
Db      1  ccaatcgtacatcgtctgtcgaatccacgctcgtcctcagagccattctgtgtgacc 60
QY      1660  GATCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1719
Db      61  aattatgtaagcagctcgtctgtaaaacacacatcgaatgcttaagaagctgttctc 120
QY      1720  TCCGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1779
Db      121  acctgatggtcattcttggtgtcgaacatctatgacatccacccgagagctcattt 180
QY      1780  TGAAGACCCCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1839
Db      181  tctagatcccttcaatctgtaaaacacatcgaatgcttaagaagctgttctc 240
QY      1840  TGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1899
Db      241  cgggtgttgggtcgtcctcaatcgtcgaagagaggtgtgtgtccctgttctcgtgga 300
QY      1900  CCAATGCTTTTCAACATCTCAAAACGATGATGATGATGATGATGATGATGATGATGATGATGAT 1959
Db      301  caagagtagctcgtcattcaccatcctcgtactgtgttgcgaagctcctcaagaacatgta 360

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 30, 2001, 22:18:33 ; Search time 2101.75 seconds
(without alignments)
12868.859 Million cell updates/sec

Title: US-08-846-234-4

Perfect score: 2517

Sequence: 1 AAAAAACACCTCTTTTA.....TTTGCAGAGTAATTAAGTGA 2517

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 537289281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estlin:*
4: em_estom:*
5: em_estopl:*
6: em_estba:*
7: em_estro:*
8: em_estrov:*
9: em_hic:*
10: gp_est1:*
11: gp_est2:*
12: gp_hic:*
13: gp_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_hiv:*
17: em_gss_pin:*
18: em_gss_pro:*
19: em_gss_rtd:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	416	16.5	668	11	BF635461 NF080A03D
2	402	16.0	649	10	AT7227515
3	371.6	14.8	649	11	BF636478
4	363.2	14.4	616	10	BE131081
5	339.6	13.5	515	10	AT1727581
6	335.4	13.3	682	11	BF634464
7	304.2	12.1	564	11	BF269676
8	295.6	11.7	489	10	AV421962
9	293.4	11.7	513	11	BF636592
10	274	10.9	818	11	BG321136
11	268	10.6	525	11	BF425566
12	266.6	10.6	530	10	AT1993310

13	264.4	10.5	435	11	BF633327
14	264	10.5	486	10	AM944715
15	251.2	10.0	591	11	BG043590
16	248.8	9.9	695	11	BF635311
17	237.8	9.4	672	11	BG454761
18	236.8	9.4	558	11	BG600526
19	236.6	9.4	435	11	BE248707
20	211.2	8.4	518	10	AM649251
21	210	8.3	597	11	BG450458
22	190.8	7.6	418	11	C72550
23	188.6	7.5	592	11	BF632552
24	187.4	7.4	540	11	BG455277
25	185.6	7.4	375	10	BE187139
26	183.6	7.3	556	11	BF632381
27	181.6	7.2	557	11	BF480406
28	181	7.2	496	10	A0083889
29	177.6	7.1	635	10	AV538951
30	174.2	6.9	411	11	T75887
31	170.4	6.8	610	11	BG648423
32	168.4	6.7	322	11	D40721
33	166.4	6.6	671	11	BG452557
34	166	6.6	307	11	BF516701
35	163.8	6.5	360	10	AT442026
36	162	6.4	446	10	BE346804
37	160.2	6.4	700	11	BI406578
38	150.2	6.0	625	10	AV541615
39	146.4	5.8	309	10	AV424385
40	143.8	5.7	571	10	AM586654
41	143.4	5.7	646	11	BG889850
42	141.2	5.6	578	10	AM585788
43	140.8	5.6	249	10	BE526884
44	140	5.6	503	10	AM155722
45	139.8	5.6	434	11	C95329

ALIGNMENTS

RESULT 1
LOCUS BF635461 668 bp mRNA EST 19-DEC-2000
DEFINITION NF080A03D|F1020 Drought Medicago truncatula cDNA clone NF080A03DT
5', mRNA sequence.
ACCESSION BF635461
VERSION BF635461.1 GI:11899619
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
Medicago.
REFERENCE 1 (bases 1 to 668)
Torrez-Jerez,I., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J.,
Flores,H.R., Imman,J.T., Weller,J.W. and May,G.D.
Expressed Sequence Tags from the Samuel Roberts Noble Foundation
Medicago truncatula drought library
Unpublished (2000)
CONTACT: May GD
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7391
Fax: 580 221 7380
Email: gdmay@noble.org
Insert Length: 668 Std Error: 0.00
Plate: 080 row: A column: 03
Seq primer: TCACACAGCAACAGCTATGAC.
FEATURES
source Location/Qualifiers
1..668
/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone="NF080A03DT"

```
/clone_lib="Drought"
/issue_type="Plantlets"
/dev_stage="Pooled timepoints"
/notes="Vector: Lambda Zap; Contains a mixture of entire
plantlets harvested in a series of days-post-watering
timepoints."
BASE COUNT      181 a      125 c      148 g      214 t
ORIGIN
Query Match      16.5%; Score 416; DB 11; Length 668;
Best Local Similarity 77.9%; Pred. No. 1.2e-83;
Matches 514; Conservative 0; Mismatches 145; Indels 1; Gaps 1;
QY 1210 GGCCTGAGAGATGTACGAGAGACTTCA-TGCTCATTTTGGAAAAAGTTGGATGAGCGGTG 1268
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8 GGTGATCAATGATGAGAGATGATCACTCTCTTTTGAAGAGATCGCAATTCATGCGGG 67
QY 1269 TTATAGATTGACGTTATTCACCTATTGAGATGTTGTGTGAAGACTATGAGAGAGATGG 1328
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 68 TCAAGTTGATGATCACTATTCATGAGATGGTGGTGAAGATGATGGAAGATTTG 127
QY 1329 ATTGGCAAGGATATTAACAAGCAATGACCAATCAATATAAATTAATTTAAAGCAA 1388
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 128 ATTTGGCCAAAGCATATTACAAAGCTCTCAACTTCAGTGAAGAAAAACATTTCAATGTA 187
QY 1389 ATGAGATCATTTGCAAGATGAAACATTTGTAAGCACTTCATTTCTTGGACGGAAGCTTA 1448
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 188 ATGCTGTCATTTGATGAGATGAAACATTTGCAATGATTTCAATGTTCTTGAACCTGAACCA 247
QY 1449 TCTCTCTTGGTCTGTTGGTGTGATGATCTTTTGGTACAGGACCCCTCGTGTGATCCAAAG 1508
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 248 TATCCCTTGGTCTGTTGGTGTGATGATTTGTGTCACCTGACCCATGATGATCCAAATG 307
QY 1509 GTACGTTTGGCTCCAAAGATGTCATGATGTCATGTCACCAAGACAGCTTGTGATGG 1568
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 308 GTATATATTTGGTCTCAAGAGATGTACATGTCATTTGTGATCAATATAGCTTATGATGAG 367
QY 1569 GGAATTCATCCACCAAGATTTGGATATGTTTCAATCTACACCCCTTGTGCTCTTTTC 427
Db 368 GAATTTTCATCCACCAAGATTTGGATATGTTTCAATCTACACCCCTTGTGCTCTTTTC 427
QY 1629 ATGCTGCTCTCGAAGCATCTCTGTGGCCGATCTATGTTAGTATCTCTGTGGGAAGC 1688
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 428 ATGAGGCTTTCAGAGCCATATCTGTGTGTCCAATTTACATTAAGGACAGCTGTGGAATC 487
QY 1689 ATACTTGTATCTCTGAAAAAACTAGTGTCTCTGATGATGATGATCTTGAAGT 1748
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 488 ACAATTTTACCTTCTCAAGAACTAGCTTGTGCTGATGATGATGATGATGATGATAA 547
QY 1749 ACTATGACTTCCGACCTGCGATTTGTTTGAAGACCTTTGCAATATGAGAAACTA 1808
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 548 CCTATGCTCCCAATGAAGACGTCTCTTCTTCTGATCTTACATGATGATGAAAACTA 607
QY 1809 TGTCTTAAATTTGAAATCTCAACAAGTTCACTGAGATGATGAGGATCACTCAAGTCCAA 1868
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 608 TGTCAAAATTTGGAACCTCAATAGTACACTGAGATTTCTGTGTGTTTAAGTCCAA 667
RESULT 2
A1727515 646 bp mRNA EST 11-JUN-1999
LOCUS BNGH18316 Six-day Cotton fiber Gossypium hirsutum cDNA 5' similar
DEFINITION to (AF073744) raffinose synthase [Cucumis sativus], mRNA sequence.
ACCESSION A1727515
VERSION A1727515.1 GI:5046367
KEYWORDS EST.
SOURCE upland cotton.
ORGANISM Gossypium hirsutum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eustosids II; Malvales; Malvaceae; Gossypium.
REFERENCE 1 (bases 1 to 646)
```

```
AUTHORS Blewitt,M., Matz,E.C., Davy,D.F. and Burr,B.
TITLE ESTs from developing cotton fiber
JOURNAL Unpublished (1999)
CONTACT Contact: Ben Burr
COMMENT Biology Department
Brookhaven National Laboratory
Upton, NY 11973, USA
Tel: 516-344-3396
Fax: 516-344-3407
Email: burr@bnl.bnl.gov
Seq primer: T3 primer.
FEATURES
source
1..646
/organism="Gossypium hirsutum"
/cultivar="Acala Maxxa"
/db_xref="taxon:3635"
/clone_lib="Six-day Cotton fiber"
/issue_type="immature fiber"
/dev_stage="Six days post anthesis"
/lab_host="XLI-Blue"
/notes="Vector: pBluescript II KS+"
BASE COUNT      162 a      139 c      149 g      193 t      3 others
ORIGIN
Query Match      16.0%; Score 402; DB 10; Length 646;
Best Local Similarity 76.3%; Pred. No. 1.7e-80;
Matches 492; Conservative 0; Mismatches 153; Indels 0; Gaps 0;
QY 1138 GCAGATGACGATGAGAGATTTGGCGGTGATTAAGATTGTTCTTCATAGTGGGCTGCT 1197
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 GAAGAAACGATGAGAGATCTTGCCTGCGATTAAGATTGTCATATGTTGTTGATGATGATG 60
QY 1198 CCGCGCGAAGAGCTGAGAGATGATGACGATTAAGAGATTCATTCATTTGAAAAATGTTG 1257
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 GCGCGCGAAGATGCTGATCAACTTATTAAGAGATTCATTCATTTGAAAAATGTTG 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1258 GATGACGATGTTTAAAGATGACGTTATTCACCTATTTGAGATGTTGTGTAAGACTATGG 1317
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 AATTGACGAGATTAAATGATGACGATTCATCTGTGGAATGCTGTGAAAACTATGG 180
QY 1318 AGGAGAGTGGATTTGGCAAGGATTAACAAGCAATGACCAATGCAATTAATAACA 1377
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 181 TGAAGAGTTGATCTTCAAAAGCTTATTAAGCTTTAAACGATTCAGTTAAAAAGCA 240
QY 1378 TTTTAAAGAAATGAGATCATGTCATATGAAAGATTTGAAGCACTTCACTGTTCTTGG 1437
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 241 TTTCAAAAGCAATGTTTATGTCAGATGAGAACCTGCAACGATTTTCATGTTCTTGG 300
QY 1438 CACGAGAGCTATCTCTGTGCTGTGTTGATGATGACTTTTGTGTGACAGACCCCTTGG 1497
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 301 AACGAAAGCCATTTGTCTGTGCTGCTGTGAGAGATTTTGTGTGCACTGATCCATCAG 360
QY 1498 TGATCAAAAGCTGTTTGGCTCAAGATGTCATATGTTATTTGTGCAACGACAG 1557
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 361 TGACCTTAATGAGCAATTTTGGCTCAAGGTTGTCAATATGTGACGTGTGTTAACAAG 420
QY 1558 CTGTGATGAGGGAATTCATCCACCCGACTGGGATGATGTTCAATCAACCCACCTTG 1617
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 421 TTTATGATGGCAACTTATACACCTGATTTGGAGATGTTCCAGTCTCCACCCCTTG 480
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1618 TGCCGCTTCATGCTGCTCTGAGCCATCTGTGTGGCCGATCATGTTAGTATTC 1677
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 481 TGTCTAGTTCCACCGCTGTTCAAGGCCATCTCGGTGGCCCAATTTTACATCANCAGAC 540
QY 1678 TGTGGAAAGCATATCTTGAATCTTGTGAAAAAATCTAGTCTTCTGATGATGATGATCCT 1737
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 541 TGTGGCAACCAACTTGTCTCTCAAAAGGCTGCTGTTTACCCGATGTTGATGATCCT 600
QY 1738 TCGAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1782
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 601 CCGTGCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 645
```

```

RESULT 3
LOCUS   BF636478
DEFINITION
5', mRNA sequence.
ACCESSION
BF636478
VERSION
BF636478.1
KEYWORDS
GI:11900636
SOURCE
barrel medic.
ORGANISM
Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
1 (bases 1 to 649)
REFERENCE
1 (bases 1 to 649)
AUTHORS
Flores-Jerez, I., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J.,
Flores, H.R., Inman, J.T., Weller, J.W. and May, G.D.
Expressed Sequence Tags from the Samuel Roberts Noble Foundation
Medicago truncatula drought library
JOURNAL
unpublished (2000)
COMMENT
Contact: May GD
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7391
Fax: 580 221 7380
Email: gdmay@noble.org
Insert Length: 649 Std Error: 0.00
Plate: 089 row: A column: 04
Seq primer: TCACACAGGAAACAGCTATGAC.
FEATURES
source
1..649
/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone_lib="NF089A04DT"
/clone_lib="Drought"
/tissue_type="plantlets"
/dev_stage="Pooled timepoints"
/Note="Vector: Lambda Zap; Contains a mixture of entire
plantlets harvested in a series of days-post-watering
timepoints."
BASE COUNT
171 a 103 c 161 g 211 t 3 others
ORIGIN
Query Match 14.8%; Score 371.6; DB 11; Length 649;
Best Local Similarity 74.1%; Pred. No. 1.2e-73;
Matches 481; Conservative 0; Mismatches 167; Indels 1; Gaps 1;

OY 1055 TGTGGATATTGGGGTGGCTTCCGCCAGAGTGGCTTGGCTTGGAGCAGCTGTGATT 1114
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 2 TGTGGTATTGGGGTGGAGAACCTAATGTTGAGGGTCTACTGAGGCTATTGTGGAG 61
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

OY 1115 CAGCCAGTCTTTCACCGAGGCTCAGATGACGATGAGGATTTGGCGGTGATTAAGTT 1174
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 62 AAGCCTAAGTTAACCGTTGGATTGGAACACTACTATGAGCATTTGGCTGTGATTAAGATT 121
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

OY 1175 GTTCTTCATTAAGTGGGGTGGTCCCGCCGAGAGAGCTGAGAGATGTACGAAGACTT 1234
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 122 GTTAATAATAGTGTGGTGTAGTTCCTCTCTCATTTTGGTTGATCAATGTATGAAGTATT 181
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

OY 1235 CATGCTCATTTGGAAAAAGTTGGATGACGGGTTAAGATTGACGTTATTCACCTATTG 1234
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 182 CACCTCTTTTGGGAAGAGTGTGAATGATGGGGTCAAGGTGATGTATTCATTCATTTGCTA 241
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

OY 1295 GAGATGTTGTGTGAAGACTATGAGAGAGATGATTTGGCAAAAGCATATTACAAGCA 1354
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 242 GAGATGTTGTGTGAAGACTATGAGAGAGATGATTTGGCCAAAGCATATTACAAGCT 301
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

OY 1355 ATGACCAATTCATTAATTAACATTTTAAAGAAATGAGCATGTGCAAGTANGGAACAT 1414
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 302 CTCTCACTTCAGTGAAGAAAAAATTCATTAATGATGATGATGCTAGACATGGAACAT 361
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

```

```

OY 1415 TGTAAAGACTTTCATGTTCCCTTGGCAGGAGACTATCTCTGTGCTGTGCTGATGAC 1474
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 362 TGCATGATTTTCATGTTGCTTGGTGAAGTGAAGCCATATCCCTTGTGTGTGATGAT 421
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

OY 1475 TTTTGGTGCAGGAGCCCTTGTGATTCGAAGGTAAGCTTTTGGCTCAAGAGTGCAC 1534
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 422 TTCTGTGTCAGTCAACCATATGTTGATCCAAATGTATGATTCATTCGAAGATGTGAC 481
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

OY 1535 ATGCTTCATTTGTGCAAGCAGACACTGTGTGATGGGGAACCTTCACCTGACTGGAT 1594
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 482 ATGTGCTATGTCATCAACATTAACCTATGATGATGGAAATTTTCACCAACGATTTGGAT 541
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

OY 1595 ATGTTCATTCACCAACCCCTTGGCGGCTTCCATGCTGCTGAGCAGCATCTGCT 1654
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 542 ATGTTCAATCTACACACCTTTGTGCTTTCATGACACTTCTANAGCCATATCTGCT 601
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

OY 1655 GGCCCGATCTATGTTAGTATCTGTGGGAAGCATTAACCTTGTATCTTC 1703
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 602 GGTCAATTTA-CATATGACACTGNTGAAATCACAATTTTGACCTTC 649
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

RESULT 4
LOCUS   BE131081
DEFINITION
L48-1031T3 Ice plant lambda Uni-Zap XR expression library, 48 hours
NaCl treatment Mesembryanthemum crystallinum cDNA clone L48-1031
5', mRNA sequence.
ACCESSION
BE131081
VERSION
BE131081.1
KEYWORDS
GI:8578444
SOURCE
common ice plant.
ORGANISM
Mesembryanthemum crystallinum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Caryophyllidae; Caryophyllales; Alzooceae; Mesembryanthemum.
1 (bases 1 to 616)
REFERENCE
Cushman, J.C.
An expressed sequence tag database for the common ice plant,
Mesembryanthemum crystallinum
unpublished (1997)
JOURNAL
Contact: Cushman JC
COMMENT
Department of Biochemistry
University of Nevada
MS200, Reno, NV 89557-0014, USA
Tel: 775-784-1918
Fax: 775-784-1650
Email: jcushman@unr.edu
PCR Primers
FORWARD: T7
REVERSE: T3
BACKWARD: T3
Plate: L48-11 row: C column: 7
Seq primer: T3
High quality sequence stop: 350
POLYA-No.
FEATURES
source
1..616
/location/Qualifiers
/organism="Mesembryanthemum crystallinum"
/db_xref="taxon:3544"
/clone_lib="L48-1031"
/clone_lib="Ice plant lambda Uni-Zap XR expression library"
/48 hours NaCl treatment"
/tissue_type="leaf, 48 h 0.4M NaCl"
/dev_stage="Six week old"
/Note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site_1:
EcoRI; Site_2: XhoI"
BASE COUNT
177 a 133 c 144 g 162 t
ORIGIN
Query Match 14.4%; Score 363.2; DB 10; Length 616;
Best Local Similarity 74.4%; Pred. No. 9.9e-72;
Matches 458; Conservative 0; Mismatches 158; Indels 0; Gaps 0;

```

QY	1458	GTCCGCTGGTGATACCTTTGGGACAGGACCCTCTCGGATGATCAAAAGGTACGTTT	1517
Db	1	GACGGCTGGGGATACCTTTTGGTGCATGCACCAATATGGAGATCCAAATGGGCACATATT	60
QY	1518	GGCTCCACGAGATGCACATATGGTTCATTTGTGCCAACAGACAGCTTGTGGATGGGAACTTCA	1577
Db	61	GGCTACAAAGTTGTCAATATATGGTCATATGTGTGCTTACAAATAGCTATATGATGGGAAATTTCA	120
QY	1578	TCCACCTGACACTGGATATGTTCCAAATCCACCACCTTTGTGCCCTTCCATGCTGCT	1637
Db	121	TCCACCTGATTTGGGATATGTTCCAAATCTACTCACTCTTGGCCAGTCTCATGCTCAT	180
QY	1638	CTCGACCACTCTCTGGTGGCCCGCATCTATGTTAGATTTCTGTGGGAAAGCATATTTG	1697
Db	181	CACGACCAATTTCCGGGTGGCCGATTTATGTAGCGATTTGTGTGGAAACACACACTTGG	240
QY	1698	ATCTTCTGAAAAAACAAGTACTGCTTCTCGATGATGATGATCTCTTGGAGTAGTACTATGAC	1757
Db	241	AGCTACTTAAAGAGTTTGTGTTTACCAGATGATCGATCATATGATAGGTGCCACATTATGAC	300
QY	1758	TCCGACCTGCGCATTTGTTGTTGTAAGACCCCTTTCATATAGGAAACATATGCTTAGA	1817
Db	301	TTCTCTACAGGATGTCCTTTGTAAGACCCCTTTCATATAGTGGAAACATATGCTTACA	360
QY	1818	TTTGAATCTCAACAAGTTCACCTGAGTAGATTTGTGCATTTCACTGCCAAAGAGAGAT	1877
Db	361	TTTGAAGACTCAACAAGTACAGAGGGGTGATTTGTGTCAATTAAGTCCAAAGAGAGAT	420
QY	1878	GGTGTGCTGAGACAGCCGCCAACCAATGCTTTTTCACAAATACTCAAAAGCATGACATCCA	1937
Db	421	GGTTCCTCAAAATGACCCCTTAACAAAGCTGCTGTGAGTTCTTCAACCCGGTAACTGCT	480
QY	1938	AACTCAACCAAAAGCATAGAAATGGACAGTAGGAGAAACCCCTATCTCTATTAAGAGCG	1997
Db	481	CAACAACCAAAAGATGTGCAATGAGCAAGGATTAAGAACCCCATCTCTCAACAAGGTG	540
QY	1998	TTTAAACCTTTGGCTTTTACCTCTATCAAGCCCAAAACCTATCTCTCAAGCCCTCTC	2057
Db	541	TTTAACTTTTTCGATTTACATTTCAAGGAGAAACCTAAGTACGCTTTAAACCCGACG	600
QY	2058	AAAGTCTTGACATATG	2073
Db	601	AAAACATTTGATTATAC	616
RESULT	5		
LOCUS	AI727581	515 bp	mRNA
DEFINITION	BNL6118471 Six-day Cotton fiber Gossypium hirsutum cDNA 5' similar to (AF073744) raffinose synthase [Cucumis sativus], mRNA sequence.		11-JUN-1999
ACCESSION	AI727581		
VERSION	AI727581.1	GI:5046433	
KEYWORDS	EST.		
ORGANISM	Gossypium hirsutum		
SOURCE	upland cotton.		
REFERENCE	Gossypium hirsutum		
AUTHORS	Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
TITLE	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;		
JOURNAL	Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.		
COMMENT	1 (bases 1 to 515)		
	Blewitt, M., Matz, E. C., Davy, D. F. and Burr, B.		
	ESTs from developing cotton fiber		
	Unpublished (1999)		
	Contact: Ben Burr		
	Biology Department		
	Brookhaven National Laboratory		
	Upton, NY 11973, USA		
	Tel: 516-344-3396		
	Fax: 516-344-3407		
	Email: burrb@nslu1.bnl.gov		
	Seq primer: T3 Primer.		
FEATURES	Location/Qualifiers		
SOURCE	1..515		

Query Match	13.5%	Score 339.6	DB 10	Length 515
Best Local Similarity	78.8%	Prod. No. 2,2e-66		
Matches 405	Conservative 0	Mismatches 109	Indels 0	Gaps 0
OY 1140	AGATGAGCATGAGAGATTGGCGGTGATAGATGTTCTTCATTAAGGTGGCGCTGCTC	1199		
Db 2	AGAAAGAGATGAGAGATCTTCCGTCGATAGATTGCAATATGTTGATGATGAGTATGCG	61		
OY 1200	CGCCGGAGAGAGCTGAGAGATGTACGAGAGACTTCATTCATTTGGAAAAATGTGGGA	1259		
Db 62	CGCCGGAGATGAGCTGATCACTTATGAAAGAAATTCATTCATTTGGAATATGTGAA	121		
OY 1260	TGCGAGGTTTAAAGATGAGCTTATCCACCATTTGAGATGTGTGAGACATATGAG	1319		
Db 122	TTGACGGAGTTAAAGTTGACGTGATTCATCTCTTGGAAATCTCTGTGAAACATATG	181		
OY 1320	GGAAGATGATTTGGCAAGAGCATATTACAAAGCAATGCCAAATCAATTAATTAACATT	1379		
Db 182	GAAAGATTGATCTTGCAAAAGCTTATTATTAAGCTTTAAACAGATAGTTAAAAAGCATT	241		
OY 1380	TTAAAGCAATGAGATCTTCAAGTATGAAACATTTGAACAGCTTCATGTTCCCTGGCA	1439		
Db 242	TCAAAGGCAATGGGTATATCCAGTATGAAACCTGCAAGATTCATGTTCTTGGAA	301		
OY 1440	CGGAAGCATATCTCTGTTGGTGGTGTGATGATGATTTGGTGACGAGACCCCTTGSTG	1499		
Db 302	CAGAAGCCATTGCTGTGGTGTGCTGTGGGAGACGATTTTGGTGACATGATCAGATG	361		
OY 1500	ATCCAAACGATGACGTTTGGCTCCAAAGGATGTCACATGTTGATTTGCCAACGACACT	1539		
Db 362	ACCTTAATGGGACATTTTGGCTCCAAAGTTGTCACATGTCACATGTCCTTAACAACATT	421		
OY 1560	TGTGGATGGGAACTTCATCCACCTGACTGCTGGATATGTTCCATTCACCCACCTTGTG	1619		
Db 422	TATGGATGGGCAACTTTATCCACCTGATTTGGACATTTCCAGTCCACACCTTGTG	481		
OY 1620	CCGCTTCACATGCTGCCCTCTGAGGCACTCTCGG	1653		
Db 482	CTGAGTTCACGCTGCTTCACAGGCGCAATCTCGG	515		

BASE COUNT

134 a 96 c 129 g 156 t

ORIGIN

/organism="Gossypium hirsutum"

/cultivar="Acala Maxxa"

/db_xref="taxon:3635"

/clone_lib="Six-day Cotton fiber"

/tissue_type="Immature fiber"

/dev_stage="Six days post anthesis"

/lab_host="XLI-Blue"

/note="Vector: pBluescript II KS+"

RESULT 6

LOCUS BE634464 662 bp mRNA EST 19-DEC-2000

DEFINITION NF061004DT1F1033 Drought Medicago truncatula cDNA clone NF061004DT

ACCESSION BE634464

VERSION BE634464.1 GI:11898622

KEYWORDS EST.

ORGANISM Medicago truncatula

SOURCE barrel medic.

REFERENCE 1 (bases 1 to 682)

AUTHORS Torres-Jerez, I., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J., Flores, H.R., Inman, J.T., Weller, J.W., and May, G.D.

TITLE Expressed Sequence Tags from the Samuel Roberts Noble Foundation Medicago truncatula drought library

JOURNAL Published (2000)

Db	124	TTTGTCGCTTTCATCGTCGCCCTCAAGACCATTTTCGTGGACCAATCTMGTCAGTGA	183
Oy	1675	TTTCGTGGGAACGATAACTTGAATCTTTCGAAAAAACTAGTCTCCTGATGCAT	1734
Db	184	CAGAATTGGGAGTCAACAACCTTGAACCTCCAGACTTGGTCTGGCTGATGGGTGCAT	243
Oy	1735	CC TTCGAAGTAGTACTATGTCACATCCGACATCCGATGGTTGTTGAAGACCCTTTGCA	1794
Db	244	CC TGAAGATGTGAGCACTATGCACTCCACCTCCAGTACGATGCTCTTTTGGCTGATCTCTCA	303
Oy	1795	TAAAGGAACACTATGCTTAAGATTGGAAATCTCAACAAAGTTCACTGAGTGAATGGTGC	1854
Db	304	TGATGGCAAACCAATGCTCAAGATATGGAACCTCAACAAGTACTGGAGTTGGGGT	363
Oy	1855	ATTCAACTGCCAAGAGGAGGATGGTGTGCTGAGACACGCCCAACCAATGCTTTTCACA	1914
Db	364	ATTTAACCTGCCAAGAGGAGGATGGTGTGCTGAGCATGATCCAAAAGGATGGCTGCA	423
Oy	1915	ATACTCAAAAGAGTGAATCCAAACCTAATCCCAAGAAGATAGATGAGCACAGTGSAGA	1974
Db	424	GTTTTCATCATCTGATCAACCAAGATGACATGCCAACACATTGATGACACAGTGTGA	483
Oy	1975	AAACCC 1980 	
Db	484	GAACCC 489	
RESULT	9		
Bf636592			
LOCUS	NF090H02D1F1027 Drought Medicago truncatula cDNA clone NF090H02DT		
DEFINITION	5', mRNA sequence.		
ACCESSION	Bf636592		
VERSION	Bf636592.1 GI:11900750		
KEYWORDS	EST.		
SOURCE	barrel medic.		
ORGANISM	Medicago truncatula		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
	Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots;		
	Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae		
	Medicago.		
REFERENCE	1 (bases 1 to 513)		
AUTHORS	Torrez-Jerez,I., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J.		
TITLE	Flores,H.R., Inman,J.T., Weller,J.W. and May,G.D.		
JOURNAL	Expressed Sequence Tags from the Samuel Roberts Noble Foundation		
COMMENT	Medicago truncatula drought library Unpublished (2000) Contact: May GD Plant Biology Division The Samuel Roberts Noble Foundation 2510 Sam Noble Parkway, Ardmore, OK 73402, USA Tel: 580 221 7391 Fax: 580 221 7380 Email: gdmay@noble.org Insert Length: 513 Std Error: 0.00 Plate: 090 row: H column: 02 Seq primer: TCACACAGGACGACGTATGAC. Location/Qualifiers 1..513 /organism="Medicago truncatula" /db_xref="taxon:3880" /clone="NF090H02DT" /clone_lib="Drought" /tissue_type="plantlets" /dev_stage="Pooled timepoints" /note="Vector: Lambda Zap; Contains a mixture of entire plantlets harvested in a series of days-post-watering timepoints."		
FEATURES			
source			
BASE COUNT	135 a 72 c 138 g 167 t 1 others		
ORIGIN			
Query Match	11.7%; score 293.4; DB 11; length 513;		

LOCUS BF425566 525 bp mRNA EST 31-JUL-2001
DEFINITION su44c11.y1 Gm-cl068 glycine max cDNA clone GENEOME SYSTEMS CLONE ID:
Gm-cl068-2061 5' similar to TR:Q92T62 Q92T62 RAFFINOSE SYNTHASE ;,
mRNA sequence.
ACCESSION BF425566
VERSION BF425566.1 GI:11413555
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
REFERENCE 1 (bases 1 to 525)
Shoemaker, R., Kelm, P., Vodkin, L., Erpelting, J., Coryell, V., Khanna
A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,
Wylie, T., Underwood, K., Steptoe, M., Rheising, B., Allen, M., Bowers
Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schuck
R., Ritzer, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann
R., Waterston, R. and Wilson, R.
Public Soybean EST Project
Unpublished (1999)
CONTACT: Shoemaker R./Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or
info@genomesystems.com web site: www.genomesystems.com
Insert length: 1160 Std Error: 0.00
High quality sequence stop: 320.
Location/Qualifiers
1. 525
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl068-2061"
/clone_1db="Gm-cl068"
/tissue_type="Leaf, drought stressed, 1 month old plants,
greenhouse grown"
/lab_host="DH10B"
/note="Vector: Bluescript II SK+; Site_1: EcoRI; Site_2:
XhoI; The cDNA library was constructed from mRNA isolated
from drought stressed leaf tissue of the cultivar Williams
82. The month old greenhouse grown plants were deprived of
water for 3 days prior to harvesting the stressed leaf
tissue. Complementary DNA was synthesized from mRNA using
a primer consisting of a poly(dT) sequence with a XhoI
restriction site. EcoRI adapters were ligated to the
blunt-ended cDNA fragments followed by XhoI digestion. The
cDNA fragments were directionally cloned into the
EcoRI-XhoI restriction site of the pBluescript vector. The
ligated cDNA fragments were transformed into DH10B host
cells (GibcoBRL). This library was constructed in the
laboratory of Dr. Randy Shoemaker."

BASE COUNT 132 a 132 c 118 g 139 t 4 others
ORIGIN

Query Match 10.6%; Score 268; DB 11; Length 525;
Best Local Similarity 69.44; Pred. No. 3.2e-50;
Matches 361; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

QY 1542 ATTGTCCACACAGCAGCTTGATGGAGGAACTTCACCCCTGATGGATATGTTCC 1601
||||| 11 ||||||| 11 ||||||| 11 ||||||| 11 ||||||| 11 |||||||
DB 2 ATTTGCGATACACAGCTTGATGGAGGAACTTCACCCACAGATTTGCCATTTCC 61
||||| 11 ||||||| 11 ||||||| 11 ||||||| 11 ||||||| 11 |||||||

QY 1602 AATCACCACACCTTTGTCGCGCTTCATGCTCCTCGAGGCCATCTCTGTTGGCCCGA 1661
||||| 11 ||||||| 11 ||||||| 11 ||||||| 11 ||||||| 11 |||||||

DB 62 AATCTACTATCTCTGTGCTGCTTCATGCTGCTCAAGACCAATCTGTGGCCCA 121
QY 1662 TCTATGTAGTATCTCTGTGCGGAAGCATATTTGATCTCTCGAAAAAAGTGTGCTTC 1721
||||| 11 ||||||| 11 ||||||| 11 ||||||| 11 ||||||| 11 |||||||
DB 122 TTTACATCATCTGACACAGCTTGGGACCAACTTTTGAAGCTTAAAGACCTTGCCCTTGC 181
QY 1722 CTGATGGATCGATCTCTCGAAGTGAAGTACTATGACACCTCCGACCTCGGATTTGTTTG 1781
||||| 11 ||||||| 11 ||||||| 11 ||||||| 11 ||||||| 11 |||||||
DB 182 CAGATGGGCTCATCTCTCAGATGTGAGCAGTATGACATCCACCCACAGGAGACTGTCTCTTG 241
QY 1782 AAGACCTTTGCATTAATGAGAAACATATGCTTAAGATTTGAAATCTACAGATTACTG 1841
||||| 11 ||||||| 11 ||||||| 11 ||||||| 11 ||||||| 11 |||||||
DB 242 CTGACCTCTCCATGATGAGCAAAACATGCTCAAGATATGAACTCAACAGTACCTG 301
QY 1842 GAGTGAATGTGCTATTAACCTGCCAAGAGAGAGATGTGCTGTGACACAGCCGCAACC 1901
||||| 11 ||||||| 11 ||||||| 11 ||||||| 11 ||||||| 11 |||||||
DB 302 GAGTCTTGTGAGCTTTTAACTGCTCAGGAGAGAGGTGCTTCNTGATTAATGTCANCA 361
QY 1902 AATGCTTTTCAATATCTCAAAACAGTGAACATCAAACTCAACCAAGACATAGAAAT 1961
||||| 11 ||||||| 11 ||||||| 11 ||||||| 11 ||||||| 11 |||||||
DB 362 AATGTGCTGCTGAGTTTCTTCATCGTGATCATCTCCAGAACATATGAAGACATTGAAT 421
QY 1962 GGCACAGTGGAGAAACCTATCTCTATGAAGGCTTAAACCTTTGCGCTTACTCT 2021
||||| 11 ||||||| 11 ||||||| 11 ||||||| 11 ||||||| 11 |||||||
DB 422 GGGATATCGAACAATAATCATGATTCAGATTCGAGGAGTCAACTTTCTCGTACGTATGA 481
QY 2022 ATCAAGCCAAAAAATTAATCTCTCTCCAGGCCCTCTCAAGA 2061
||||| 11 ||||||| 11 ||||||| 11 ||||||| 11 ||||||| 11 |||||||
DB 482 ACCAAGCCCAAGAACATCATCTCTCGGCGCATTTGATGA 521

RESULT 12
LOCUS AT1993310 530 bp mRNA EST 08-SEP-1999
DEFINITION 701496022 A. thaliana, Ohio State clone set Arabidopsis thaliana
cDNA clone 701496022, mRNA sequence.
ACCESSION AT1993310
VERSION AT1993310.1 GI:5840215
KEYWORDS EST.
SOURCE EST.
ORGANISM Arabidopsis thaliana
thale cress.
REFERENCE 1 (bases 1 to 530)
Chen, J., Momiyama, M., Chan, E., Mooney, M., Carroon, B., Gilliland, D.,
Wang, X., Hillman, J., Guegler, K., Kim, C., Doyle, M., Brzoska, P.,
Gorgone, G., Burns, D., Griffin, J., Mouanoutoua, M., Nguyen, D., Tan, R.,
Rose, M., Warren, B., Ton, B., Kastury, K., Borillo, C., Carpio, T.,
Policky, J., Suzuki, G., Argentine, C., Shah, S., Nobrigha, A., Murry, L.,
Turner, C., Krikorian, S., Elder, L. and Hanson, D.
Arabidopsis thaliana Gene Expression Microarray
Unpublished (1999)
CONTACT: David Smoller, Ph.D.
Genome Systems, Inc., a wholly owned subsidiary of Incyte
Pharmaceuticals, Inc.
4633 World Parkway Circle, St. Louis, MO 63134, USA
Tel: 877-577-2733
Fax: 314-427-3324
Email: service@genomesystems.com.
Location/Qualifiers
1. 530
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/clone="701496022"
/clone_1db="A. thaliana, Ohio State clone set"
/note="cDNA library was made from selected clones from the
Arabidopsis thaliana Ohio State clone set."

BASE COUNT 132 a 115 c 128 g 155 t
ORIGIN

Query Match 10.6%; Score 266.6; DB 10; Length 530;

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 1, 2001, 00:44:50 ; Search time 86.51 Seconds
(without alignments)
6589.355 Million cell updates/sec

Title: US-08-846-234-4

Perfect score: 2517

Sequence: 1 AAAAAACACCCCTCTTTTA.....TTTGAGAGATTAAAGTGA 2517

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_patents_NA:*
1: /cgn2_6/ptodata/2/1na/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/1na/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/1na/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/1na/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/1na/PCrUS.COMB.seq:*
6: /cgn2_6/ptodata/2/1na/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2517	100.0	2517	4	Sequence 4, Appl1
2	47.4	1.9	7218	1	Sequence 14, Appl1
3	40.8	1.6	7218	1	Sequence 14, Appl1
4	40.4	1.6	1631	3	Sequence 1, Appl1
5	38	1.5	911	4	Sequence 19, Appl1
6	36.4	1.4	5511	3	Sequence 21, Appl1
7	36.4	1.4	7334	3	Sequence 2, Appl1
8	36	1.4	2824	2	Sequence 3, Appl1
9	36	1.4	8083	4	Sequence 4, Appl1
10	36	1.4	8083	4	Sequence 5, Appl1
11	35.6	1.4	2793	1	Sequence 1, Appl1
12	35.6	1.4	2793	1	Sequence 1, Appl1
13	35.2	1.4	2338	2	Sequence 1, Appl1
14	35.2	1.4	2338	2	Sequence 1, Appl1
15	35.2	1.4	1326	4	Sequence 53, Appl1
16	34.8	1.4	18596	4	Sequence 11, Appl1
17	34.8	1.4	289	4	Sequence 17, Appl1
18	34.6	1.4	289	4	Sequence 17, Appl1
19	34.6	1.4	216	1	Sequence 45, Appl1
20	34.4	1.4	216	4	Sequence 45, Appl1
21	34.4	1.4	216	4	Sequence 45, Appl1
22	34.4	1.4	2032	5	Sequence 5, Appl1
23	34.4	1.3	1416	4	Sequence 3, Appl1
24	33.6	1.3	2509	2	Sequence 1, Appl1
25	33.6	1.3	4377	3	Sequence 28, Appl1
26	33.6	1.3	4377	3	Sequence 28, Appl1
27	33.6	1.3	4377	3	Sequence 28, Appl1

C	28	33.6	1.3	4377	4	US-09-479-409-28	Sequence 28, Appl1
C	29	33.4	1.3	1505	1	US-07-915-246-1	Sequence 1, Appl1
C	30	33.4	1.3	16442	3	US-08-781-891-208	Sequence 208, Appl1
C	31	33	1.3	34303	2	US-08-735-609-4	Sequence 4, Appl1
C	32	33	1.3	34303	2	US-08-735-609-4	Sequence 4, Appl1
C	33	33	1.3	34303	3	US-09-315-372-4	Sequence 4, Appl1
C	34	33	1.3	34303	3	US-09-244-752-4	Sequence 4, Appl1
C	35	33	1.3	34303	3	US-09-245-497-4	Sequence 4, Appl1
C	36	33	1.3	34382	2	US-08-374-483-6	Sequence 6, Appl1
C	37	33	1.3	35408	4	US-08-973-334-3	Sequence 3, Appl1
C	38	33	1.3	35408	4	US-09-563-869A-3	Sequence 3, Appl1
C	39	33	1.3	35408	4	US-08-549-489-3	Sequence 3, Appl1
C	40	33	1.3	35935	2	US-08-735-609-1	Sequence 1, Appl1
C	41	33	1.3	35935	2	US-08-735-609-1	Sequence 1, Appl1
C	42	33	1.3	35935	3	US-08-379-452-43	Sequence 43, Appl1
C	43	33	1.3	35935	3	US-09-315-372-1	Sequence 1, Appl1
C	44	33	1.3	35935	3	US-09-244-752-1	Sequence 1, Appl1
C	45	33	1.3	35935	3	US-09-245-497-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1
US-08-846-234-4
; Sequence 4, Application US/08846234
; Patent No. 6166292
; GENERAL INFORMATION:
; APPLICANT: OSUMI Chieko
; APPLICANT: NOZAKI Jinshi
; APPLICANT: KIDA Takao
; TITLE OF INVENTION: RAFFINOSE SYNTHASE GENE, METHOD FOR
; TITLE OF INVENTION: PRODUCING RAFFINOSE, AND TRANSGENIC PLANT
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESS: OBLON, SPIVAK, MCCLELLAND, MATIER & NEUSTADT, P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,234
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)-413-3000
; TELEFAX: (703)-413-2220
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2517 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ORIGINAL SOURCE:
; ORGANISM: cucumber (Cucumis sativus)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 56..2407
; US-08-846-234-4
Query Match 100.0%; Score 2517; DB 4; Length 2517;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2517; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 AAAAAACACCTCTCTTATGTTTTTGGGTTGTTCTCTCTTCTCTCAAAATGSC 60
1 AAAAAACACCTCTCTTATGTTTTTGGGTTGTTCTCTCTTCTCTCAAAATGSC 60
61 TCTTGTGTTTTAAAAATGCTGGCTCCACAGTACTTTCATTGATGGCTTAAATGACATGTC 120
61 TCTTGTGTTTTAAAAATGCTGGCTCCACAGTACTTTCATTGATGGCTTAAATGACATGTC 120
61 TCTTGTGTTTTAAAAATGCTGGCTCCACAGTACTTTCATTGATGGCTTAAATGACATGTC 120
121 GTACCGGTTTGCATTCGACGATCGATCGATTCACGTGACGAGGTCATTCCTCTGTCCGA 180
121 GTACCGGTTTGCATTCGACGATCGATCGATTCACGTGACGAGGTCATTCCTCTGTCCGA 180
121 GTACCGGTTTGCATTCGACGATCGATCGATTCACGTGACGAGGTCATTCCTCTGTCCGA 180
181 TGTTCCTGAGAACATTTGTTGCTTCCTCTCTCCGTAACACTTCGATGACAGTCCCGGT 240
181 TGTTCCTGAGAACATTTGTTGCTTCCTCTCTCCGTAACACTTCGATGACAGTCCCGGT 240
181 TGTTCCTGAGAACATTTGTTGCTTCCTCTCTCCGTAACACTTCGATGACAGTCCCGGT 240
241 TTTGCGTGTGCTGCTTTGTTGATTCGACCGCGTCGAACTGATAGCCGATGTTTCTTTC 300
241 TTTGCGTGTGCTGCTTTGTTGATTCGACCGCGTCGAACTGATAGCCGATGTTTCTTTC 300
241 TTTGCGTGTGCTGCTTTGTTGATTCGACCGCGTCGAACTGATAGCCGATGTTTCTTTC 300
301 GATTGGGAGCTGAAGGATATTGCTTTATGAGTATTTTCAGGTTTAAAGTTTGTGTCAC 360
301 GATTGGGAGCTGAAGGATATTGCTTTATGAGTATTTTCAGGTTTAAAGTTTGTGTCAC 360
301 GATTGGGAGCTGAAGGATATTGCTTTATGAGTATTTTCAGGTTTAAAGTTTGTGTCAC 360
361 TACACACTGGGTTGTCGAAATGCTGGGATCTTGAATCGAGACTCAGATTGTGATCCT 420
361 TACACACTGGGTTGTCGAAATGCTGGGATCTTGAATCGAGACTCAGATTGTGATCCT 420
361 TACACACTGGGTTGTCGAAATGCTGGGATCTTGAATCGAGACTCAGATTGTGATCCT 420
421 TGTGAGTGAAGTATTCGTGTCGACCGGATGTTTCTCTTCGATGATGTTGAGGACCGTT 480
421 TGTGAGTGAAGTATTCGTGTCGACCGGATGTTTCTCTTCGATGATGTTGAGGACCGTT 480
421 TGTGAGTGAAGTATTCGTGTCGACCGGATGTTTCTCTTCGATGATGTTGAGGACCGTT 480
481 CCGAACCCTGATTCAGCTCGACCGGATGATGATGATGATGATGATGATGATGATGATGATG 540
481 CCGAACCCTGATTCAGCTCGACCGGATGATGATGATGATGATGATGATGATGATGATGATG 540
481 CCGAACCCTGATTCAGCTCGACCGGATGATGATGATGATGATGATGATGATGATGATGATG 540
541 GTTCGAAAGTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600
541 GTTCGAAAGTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600
541 GTTCGAAAGTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600
601 GTTTCGACTGTTTAAAGAGGCGATGAGATGATGAGATGATGAGATGATGAGATGATGAGATG 660
601 GTTTCGACTGTTTAAAGAGGCGATGAGATGATGAGATGATGAGATGATGAGATGATGAGATG 660
601 GTTTCGACTGTTTAAAGAGGCGATGAGATGATGAGATGATGAGATGATGAGATGATGAGATG 660
661 GTTTCGAGGAGAACTCCACAGGATATGCTGACCAAAATTCGTTGTTGTCAGTGGGAGCG 720
661 GTTTCGAGGAGAACTCCACAGGATATGCTGACCAAAATTCGTTGTTGTCAGTGGGAGCG 720
661 GTTTCGAGGAGAACTCCACAGGATATGCTGACCAAAATTCGTTGTTGTCAGTGGGAGCG 720
721 GTTTCGAGGAGAACTCCACAGGATATGCTGACCAAAATTCGTTGTTGTCAGTGGGAGCG 780
721 GTTTCGAGGAGAACTCCACAGGATATGCTGACCAAAATTCGTTGTTGTCAGTGGGAGCG 780
721 GTTTCGAGGAGAACTCCACAGGATATGCTGACCAAAATTCGTTGTTGTCAGTGGGAGCG 780
781 CGGTTGCTCCCGGTTTAAAGAGGCGATGAGATGATGAGATGATGAGATGATGAGATGATGAGATG 840
781 CGGTTGCTCCCGGTTTAAAGAGGCGATGAGATGATGAGATGATGAGATGATGAGATGATGAGATG 840
781 CGGTTGCTCCCGGTTTAAAGAGGCGATGAGATGATGAGATGATGAGATGATGAGATGATGAGATG 840
841 GATTCCTCATCACCAGAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 900
841 GATTCCTCATCACCAGAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 900
841 GATTCCTCATCACCAGAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 900
901 TCTTTTGAATTCACAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 960
901 TCTTTTGAATTCACAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 960
901 TCTTTTGAATTCACAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 960
961 CCCCCGAGCCGCGCAGAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 1020
961 CCCCCGAGCCGCGCAGAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 1020
961 CCCCCGAGCCGCGCAGAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 1020
1021 GAGTGTGAGACATGTTTATGTTTGGATGCTTGTGTGATGATGATGATGATGATGATGATGATGATGATGATG 1080
1021 GAGTGTGAGACATGTTTATGTTTGGATGCTTGTGTGATGATGATGATGATGATGATGATGATGATGATGATG 1080

1081 GGAGGTGCTGGCTGCTGAGGACGTGATTCAGCCAGTGTCTTACCAAGGGCTGCA 1140
1081 GGAGGTGCTGGCTGCTGAGGACGTGATTCAGCCAGTGTCTTACCAAGGGCTGCA 1140
1141 GATGACGATGAGAGATTTGGCGGTGATAGATGTTCTTCATTAAGTGGCGTGTCTCC 1200
1141 GATGACGATGAGAGATTTGGCGGTGATAGATGTTCTTCATTAAGTGGCGTGTCTCC 1200
1141 GATGACGATGAGAGATTTGGCGGTGATAGATGTTCTTCATTAAGTGGCGTGTCTCC 1200
1201 GCCGGAAGGCTGAGGAGATGACGAAGGCTTCATGCTCANTTGGAAAAAGTTGGGAT 1260
1201 GCCGGAAGGCTGAGGAGATGACGAAGGCTTCATGCTCANTTGGAAAAAGTTGGGAT 1260
1201 GCCGGAAGGCTGAGGAGATGACGAAGGCTTCATGCTCANTTGGAAAAAGTTGGGAT 1260
1261 CGACGGTGTAAAGTATGAGCTTATCCACTATGAGATGTTGTGAGAGCTATGAGAG 1320
1261 CGACGGTGTAAAGTATGAGCTTATCCACTATGAGATGTTGTGAGAGCTATGAGAG 1320
1261 CGACGGTGTAAAGTATGAGCTTATCCACTATGAGATGTTGTGAGAGCTATGAGAG 1320
1321 GAGAGTGAATTTGGCAAGGATATTTACAAAGCATGACCAAAATCAATTAATTAACATTT 1380
1321 GAGAGTGAATTTGGCAAGGATATTTACAAAGCATGACCAAAATCAATTAATTAACATTT 1380
1321 GAGAGTGAATTTGGCAAGGATATTTACAAAGCATGACCAAAATCAATTAATTAACATTT 1380
1381 TAAAGGAATGAGTCAATTCAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 1440
1381 TAAAGGAATGAGTCAATTCAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 1440
1381 TAAAGGAATGAGTCAATTCAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 1440
1441 GGAAGCTATCTCTGTCGTCGTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1500
1441 GGAAGCTATCTCTGTCGTCGTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1500
1441 GGAAGCTATCTCTGTCGTCGTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1500
1501 TCCAAACGCTATCTTTGCTGCTCAAGAGATGACATGATGATGATGATGATGATGATGATGATGATGATG 1560
1501 TCCAAACGCTATCTTTGCTGCTCAAGAGATGACATGATGATGATGATGATGATGATGATGATGATGATG 1560
1501 TCCAAACGCTATCTTTGCTGCTCAAGAGATGACATGATGATGATGATGATGATGATGATGATGATGATG 1560
1561 GTGAGTGGGAACTTCAATCCACCGTGAAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 1620
1561 GTGAGTGGGAACTTCAATCCACCGTGAAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 1620
1561 GTGAGTGGGAACTTCAATCCACCGTGAAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 1620
1621 CGCCTTCATGCTGCTCTGTCGACCATCTCTGTTGGCCGATCTATGTTAGTATCTGT 1680
1621 CGCCTTCATGCTGCTCTGTCGACCATCTCTGTTGGCCGATCTATGTTAGTATCTGT 1680
1621 CGCCTTCATGCTGCTCTGTCGACCATCTCTGTTGGCCGATCTATGTTAGTATCTGT 1680
1681 GGGAAAGCATTAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1740
1681 GGGAAAGCATTAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1740
1681 GGGAAAGCATTAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1740
1741 AAGTGAATG 1800
1741 AAGTGAATG 1800
1741 AAGTGAATG 1800
1801 AGAAATG 1860
1801 AGAAATG 1860
1801 AGAAATG 1860
1861 CTGCGAAGAGAGAGATG 1920
1861 CTGCGAAGAGAGAGATG 1920
1861 CTGCGAAGAGAGAGATG 1920
1921 AAAACGATG 1980
1921 AAAACGATG 1980
1921 AAAACGATG 1980
1981 TATCTTATGAGAGGCTTAAACCTTTGGCTTTTACCTGATCAAGCCAAAATTTAT 2040
1981 TATCTTATGAGAGGCTTAAACCTTTGGCTTTTACCTGATCAAGCCAAAATTTAT 2040
1981 TATCTTATGAGAGGCTTAAACCTTTGGCTTTTACCTGATCAAGCCAAAATTTAT 2040
2041 CCTCTCCAAAGCCCTCTCAAGATCTTGAATAGCTTTGACCATTTGCAATTTGAGCTCAT 2100
2041 CCTCTCCAAAGCCCTCTCAAGATCTTGAATAGCTTTGACCATTTGCAATTTGAGCTCAT 2100
2041 CCTCTCCAAAGCCCTCTCAAGATCTTGAATAGCTTTGACCATTTGCAATTTGAGCTCAT 2100
2101 CACTGTTTACCAAGTGAACCAATCTCAATCTTCAATCTTCAATCTTCAATCTTCAATCTTCAATCTTCAATCTT 2160
2101 CACTGTTTACCAAGTGAACCAATCTCAATCTTCAATCTTCAATCTTCAATCTTCAATCTTCAATCTTCAATCTT 2160
2101 CACTGTTTACCAAGTGAACCAATCTCAATCTTCAATCTTCAATCTTCAATCTTCAATCTTCAATCTTCAATCTT 2160
2161 GGTGAACATGTTTAAACTAGTAGGACCATTCATCTGTGAGTATGACATGACATTAAG 2220

```
|||||
Db 2161 GGTGAACATGCTTAACAGTGGAGGACCATCCATCTGTGACATATGACGATGACCTAG 2220
Qy 2221 CTCAGTCGAGATGGTGTCAAAAGGCTGTGGTGAGATGCGAGTATTTGCATCGAAAAACC 2280
Db 2221 CTCAGTCGAGATGGTGTCAAAAGGCTGTGGTGAGATGCGAGTATTTGCATCGAAAAACC 2280
Qy 2281 AAGGCTGTGCTATGATGGGAGAGATGTTGGTTCAGTATGATGATGAGCAAAATGCT 2340
Db 2281 AAGGCTGTGCTATGATGGGAGAGATGTTGGTTCAGTATGATGATGAGCAAAATGCT 2340
Qy 2341 GGTGCTCAAGTCCGCAATGATTCCTTCATCGGCTGCGATTTGCGTATGAGTA 2400
Db 2341 GGTGCTCAAGTCCGCAATGATTCCTTCATCGGCTGCGATTTGCGTATGAGTA 2400
Qy 2401 CTGTTTAAATTTTATTTATGTAAGCTCAATGATTTGTTGCTGCTGCTGCT 2460
Db 2401 CTGTTTAAATTTTATTTATGTAAGCTCAATGATTTGTTGCTGCTGCTGCTGCT 2460
Qy 2461 ATCAATGATTTCTCTCCAAAGAAATATGTAATTTGAGAGATTAATGA 2517
Db 2461 ATCAATGATTTCTCTCCAAAGAAATATGTAATTTGAGAGATTAATGA 2517
```

RESULT 2

```
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
```

GENERAL INFORMATION:

```
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESS: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300, 6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
```

INFORMATION FOR SEQ ID NO: 14:

```
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: PTZ9PT-F18
US-08-232-463-14
```

Query Match 1.9%; Score 47.4; DB 1; Length 7218;
Best Local Similarity 4.7%; Pred. No. 0.0005;
Matches 15; Conservative 178; Mismatches 124; Indels 0; Gaps 0;

```
Qy 1132 AGGCTGACGATGAGATGAGAGATTTGGCGGTGATAGATTTCTTCATAGGCTCG 1191
Db 1356 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1297
Qy 1192 GCTGTCGCCGCGGAGAAAGCTGAGAGATGTACGAGACTTCATGCTCATTTGAAA 1251
Db 1296 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1237
Qy 1252 AGTTGGATCGACGCTGTAAATGATGAGCTTATCCACCATTTGAGAGATGTTGTGA 1311
Db 1236 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1177
Qy 1312 CTATGAGAGAGATGATTTGCAAGCATATTACAAAGCAATGACCAATCATATA 1371
Db 1176 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1117
Qy 1372 TAAACATTTAAAGAAATGAGCTCATTTGCAAGTATGAAACATTTGACGACTTCAT 1431
Db 1116 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1057
Qy 1432 CCTGGCAGCAAGCTA 1448
Db 1056 CCTGACCTGCAGCA 1040
```

RESULT 3

```
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
```

GENERAL INFORMATION:

```
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESS: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300, 6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
```

INFORMATION FOR SEQ ID NO: 14:

```
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
```

STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: PF2gpt-Fls
US-08-232-463-14

Query Match 1.6%; Score 40.8; DB 1; Length 7218;
Best Local Similarity 2.2%; Pred. No. 0.057;
Matches 6; Conservative 163; Mismatches 105; Indels 0; Gaps 0;

QY 10 CCCTTCTTTAGTTTGTGGGTTGTTCTCTCTTCTCTCTCAAAATGGCTCTAGTTT 69
DB 1170 YY 1229
QY 70 TAAAAATGGGCGCTCCAGCTAGTTCAATTGAGTTAAATGACATGCTGACCGTT 129
DB 1230 YY 1289
QY 130 TGCATCGACGATCGGATTCAGTGTGAACGGTCATTCGTTCTGCGATGTTCTGA 189
DB 1290 YY 1349
QY 190 GAACATGTTGCTTCCTCTCTCCGTAACACTGATAGACAACTCCCGGTTGCGTTG 249
DB 1350 YY 1409
QY 250 TTGCTTTGTTGATTCGACGCGTCCGAACTGAT 283
DB 1410 YYYYYYYYYYYYYYYYYYYYYYYYYYGTACCAAT 1443

RESULT 4
US-09-118-319-1/c
; Sequence 1, Application US/09118319
; Patent No. 6114158
; GENERAL INFORMATION:
; APPLICANT: Li, Xin-Liang
; APPLICANT: Chen, Huizhong
; APPLICANT: Ljungdahl, Lars G.
; TITLE OF INVENTION: Orpinomyces Cellulase Celf Protein and Coding Sequences
; FILE REFERENCE: 33-98sequence listing
; CURRENT APPLICATION NUMBER: US/09/118,319
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1631
; TYPE: DNA
; ORGANISM: Orpinomyces sp. PC-2
; FEATURE:
; NAME/KEY: Intron
; LOCATION: (187)..(297)
US-09-118-319-1

Query Match 1.6%; Score 40.4; DB 3; Length 1631;
Best Local Similarity 65.6%; Pred. No. 0.028;
Matches 59; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 2396 GAGTACTGTTTAAATTTTATTTATGTAAGCTCAATGATGTTGTTGCTGTTG 2455
DB 501 GAGTAAATGTTGTTG 442
QY 2456 TTGCTATCAATGATTTCTCTCCAAAGAA 2485
DB 441 TTGTAATCAATGATTTATTTGTTGAAGCA 412

RESULT 5
US-09-457-046B-19
; Sequence 19, Application US/09457046B
; Patent No. 6287835

GENERAL INFORMATION:
APPLICANT: Croteau, Rodney et al.
TITLE OF INVENTION: Transacylases of the Paclitaxel Biosynthetic Pathway
FILE REFERENCE: 53679
CURRENT APPLICATION NUMBER: US/09/457,046B
CURRENT FILING DATE: 1999-12-07
NUMBER OF SEQ ID NOS: 74
SOFTWARE: Patentln Ver. 2.1
SEQ ID NO 19
LENGTH: 911
TYPE: DNA
ORGANISM: Taxus cuspidata
US-09-457-046B-19

Query Match 1.5%; Score 38; DB 4; Length 911;
Best Local Similarity 47.2%; Pred. No. 0.11;
Matches 116; Conservative 0; Mismatches 130; Indels 0; Gaps 0;

QY 372 TTGTCGAATGTTGGGATCTTGATCGGAGACTCAGATTGATCTTGAGAACTCAG 431
DB 461 ttgtataaatgltgacacacataagataatgaacaatgltatggaagaaatgtaag 520
QY 432 ATTCTGTCGACCGTATGTTTCTCTTCCGATCGTTGAGGACCGTTCCGAACTCGA 491
DB 521 attttgttcgtcccttgaaagtagtgagcagcatglttgatagaagaacaagcctc 580
QY 492 TTCAGCCGCGGATGATGACTTTGTCATGTTGTGTCGAGAGTGTGCTCGAAAGTTG 551
DB 581 tcaaatccacatactagaaatgaaagcttcttcgcatggaatggaagaaatlat 640
QY 552 TTGATGATGTCCTCCGAAGTATGTTGATCTCATGCTGATGATCCGTTGCACTTG 611
DB 641 ttaatccacatactagaaatgaaagcttcttcgcatggaatggaagaaatlat 700
QY 612 TTAAG 617
DB 701 ataag 706

RESULT 6
US-09-457-046B-21
; Sequence 21, Application US/09457046B
; Patent No. 6287835
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney et al.
; TITLE OF INVENTION: Transacylases of the Paclitaxel Biosynthetic Pathway
; FILE REFERENCE: 53679
; CURRENT APPLICATION NUMBER: US/09/457,046B
; CURRENT FILING DATE: 1999-12-07
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 21
; LENGTH: 911
; TYPE: DNA
; ORGANISM: Taxus cuspidata
US-09-457-046B-21

Query Match 1.5%; Score 38; DB 4; Length 911;
Best Local Similarity 47.2%; Pred. No. 0.11;
Matches 116; Conservative 0; Mismatches 130; Indels 0; Gaps 0;

QY 372 TTGTCGAATGTTGGGATCTTGATCGGAGACTCAGATTGATCTTGAGAACTCAG 431
DB 461 ttgtataaatgltgacacacataagataatgaacaatgltatggaagaaatgtaag 520
QY 432 ATTCTGTCGACCGTATGTTTCTCTTCCGATCGTTGAGGACCGTTCCGAACTCGA 491
DB 521 attttgttcgtcccttgaaagtagtgagcagcatglttgatagaagaacaagcctc 580
QY 492 TTCAGCCGCGGATGATGACTTTGTCATGTTGTGTCGAGAGTGTGCTCGAAAGTTG 551

APPLICATION NUMBER: <unknown>
FILING DATE: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Friedman, Mark M.
REGISTRATION NUMBER: 33,883
REFERENCE/DOCKET NUMBER: 1402/2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 972-3-562553
TELEFAX: 972-3-562554
TELEX: <unknown>
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 8083
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-383-630-5

Query Match 1.4%; Score 36; DB 4; Length 8083;
Best Local Similarity 50.6%; Pred. No. 1.9;
Matches 87; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 797 TTATGCTAATGAGAGATGTTGGCAATCCATGAGACGATTCGATCCATCACCANA 856
DB 4977 TTTCCTTCAACCGGAGAGCTGTGGAAAAAAGGAGACGATATGATCATCATCAAG 4918
QY 857 GAAGAATGACCAACCGCTCGCGGAGCAAAATGCGCTGTTTGAATTCCAA 916
DB 4917 GAACCTGAGACTAGCCGCGGGGATCTGAGCCCGGGGATCATCAAGATCCCT 4858
QY 917 GAGATTACAAATTCGCTGATCACTCAATCCCAAGCCAGCCGCCGAG 968
DB 4857 AAGGTTACTGTCTCCACACTAAGCACACAGAGGCTTGCTCACCAG 4806

RESULT 12
US-08-209-747-1
Sequence 1, Application US/08209747
Patent No. 5733771
GENERAL INFORMATION:
APPLICANT: Lewis, Randolph V.
TITLE OF INVENTION: CDNA Encoding Minor Ampullate Spider
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: P.O. Box 747
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22040-3487
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/209,747
FILING DATE: 14-MAR-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1447-104P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-205-8000
TELEFAX: 703-205-8050
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2793 base pairs

TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Nephila clavipes
TISSUE TYPE: minor ampullate gland
FEATURE:
NAME/KEY: CDS
LOCATION: 183..2675
OTHER INFORMATION: /product= "N. clavipes minor
OTHER INFORMATION: ampullate silk protein"
US-08-209-747-1

Query Match 1.4%; Score 35.6; DB 1; Length 2793;
Best Local Similarity 52.7%; Pred. No. 1.3;
Matches 77; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 1084 GGTGCTGCTTCCCTGAGGACGCTGTGATTCAGCCAGCTTTTACCCAGGCGTCAGAT 1143
DB 1679 GGTGCGGAGAGAGAGCTGTGCTGCTGCTGTGAGAGAGAGAGAGCTGTGT 1738
QY 1144 GACGATGAGAGATTTGGCGGTGATATGATTTCTTCAATAGTGTGGGCTGTCGCC 1203
DB 1739 TACGATGAGAGTGTGTGCTGTGCTGAGCTGCTGTGAGGTGTGAGCTGTGAGCCCT 1798
QY 1204 GGAGAGGCTGAGAGATGTACGAG 1229
DB 1799 GCAGGTGACGAGAGAGAGCTGGAG 1824

RESULT 13
US-08-458-298-1
Sequence 1, Application US/08458298
Patent No. 5756677
GENERAL INFORMATION:
APPLICANT: Lewis, Randolph V.
TITLE OF INVENTION: CDNA Encoding Minor Ampullate Spider
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: P.O. Box 747
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22040-3487
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/458,298
FILING DATE: 02-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/209,747
FILING DATE: 14-MAR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1447-104P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-205-8000
TELEFAX: 703-205-8050
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2793 base pairs
TYPE: nucleic acid

```

? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: CDNA
? HYPOTHETICAL: NO
? ORIGINAL SOURCE:
? ORGANISM: Nephila clavipes
? TISSUE TYPE: minor ampullate gland
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 183..2675
? OTHER INFORMATION: /product= "N. clavipes minor
? ampullate silk protein"
? OS-08-456-298-1

```

Query Match	1.48;	Score 35.6;	DB 1;	Length 2793;
Best Local Similarity	52.7%;	Pred. No. 1.3;		
Matches 77;	Conservative	0;	Mismatches 69;	Indels 0;
			Gaps	0;

Accession	Sequence	Position
QY 1084	GGCCCTGGCTTCCCTGAGGACAGTGTGATTTAGCAGCAGTTCACACAGGGCTCAAT	1143
Db 1679	GGTCCGGACACAGACTGGTGGCGCTGCTGCTGGTGGTGCAGAGACAGAGAGTCTGGT	1738
QY 1144	GAGCATGAGAGATTTGGCGGTGGATTAAGTTTCTCTCAATAAGTCCGGGCTGGTCCGCC	1203
Db 1739	TACGGTATGAGAGTGTGGTGGTCTGGAGCTGCTGCAGGTGCTGGAGCTGGAGCCCT	1788
QY 1204	GGAGAGGCTGAGAGAGTGTTCGAAC	1229
Db 1799	GCAGGTGCAGAGCAGAGAGCTGGAGG	1824

RESULT 14
US-08-425-069-1

```

TOPOLGY: linear
MOLECULE TYPE: CDNA
HYPOTHEetical: NO
ORIGINAL SOURCE:
ORGANISM: Nephilia clavipes
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2154
OTHER INFORMATION: /product= "Nephilia clavipes
OTHER INFORMATION: dragline silk protein"
PUBLICATION INFORMATION:
AUTHORS: Xu, Ming
AUTHORS: Lewis, Randolph V.
TITLE: Structure of a protein superfiber: Spider
TITLE: draftline silk
JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
VOLUME: 87
PAGES: 7120-7124
DATE: Sept.-1990
RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 2338
IS-08-425-069-1

```

Query Match	1.4%;	Score 35.2;	DB 1;	Length 2338;
Best Local Similarity	49.0%;	Pred. No. 1.5;		
Matches	94;	Conservative	0;	Mismatches 98;
				Indels 0;
				Gaps 0;

Accession	Sequence	Position
Oy	1083 AGATGCGCTGGCTGCGCTGAGGACACGGTGTATTCACGCAATGCTTTACACAGGGCTGCAGA	1142
Oy	1281 AGGTGGATTAAAGTGTGACAAAGGGGCGAGGTGACGACGCCGCTGCACGACGCTGGAGGTGCGCG	1340
Oy	1143 TGACGATGAGAGAGATTGGCGGTGAGTAAAGATTGTTCTCATAAAGGTGCGGGCTGCTGCCCG	1202
Db	1341 ACAAGAGAGGATATGAGACGCTTGGAAACCAAGTGCCTGGACAGCAGTGAGCAAGGTGCAGC	1400
Oy	1203 CGGAGAAAGGCTGAGGAGATGTACGAAGACCTTCATTTGGAAAAAGTTGGGATCG	1262
Db	1401 AGCAGCAGCTGGAGAGTCCCGACAAAGAGATATGAGAGGCTTTGGAAAGCCAAAGGTGCTGG	1460
Oy	1263 ACGGTATTAAAGA	1274
Db	1461 ACGAGGTGACAA	1472

RESULT 15
US-08-317-844B-1
; Sequence 1, Application US/08317844B
; Patent No. 5989894
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V.
; APPLICANT: Xu, Ming
; APPLICANT: Himman, Michael B.
; TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
; TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
; TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
; NUMBER OF SEQUENCES: 62
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: 301 No. 5989894th Washington Street
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22046
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/317,844B
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 30, 2001, 09:42:57 : Search time 70.16 Seconds
(without alignments)
827.729 Million cell updates/sec

Title: US-08-846-234-5
Perfect score: 4225
Sequence: 1 MAPSFKNKGNVWSEFDGLND.....QVWPIDSSSGISVIEYLF 784

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_1101: *
1: /SIDSR/gcgdata/geneseq/geneseq/AA1980.DAT: *
2: /SIDSR/gcgdata/geneseq/geneseq/AA1981.DAT: *
3: /SIDSR/gcgdata/geneseq/geneseq/AA1982.DAT: *
4: /SIDSR/gcgdata/geneseq/geneseq/AA1983.DAT: *
5: /SIDSR/gcgdata/geneseq/geneseq/AA1984.DAT: *
6: /SIDSR/gcgdata/geneseq/geneseq/AA1985.DAT: *
7: /SIDSR/gcgdata/geneseq/geneseq/AA1986.DAT: *
8: /SIDSR/gcgdata/geneseq/geneseq/AA1987.DAT: *
9: /SIDSR/gcgdata/geneseq/geneseq/AA1988.DAT: *
10: /SIDSR/gcgdata/geneseq/geneseq/AA1989.DAT: *
11: /SIDSR/gcgdata/geneseq/geneseq/AA1990.DAT: *
12: /SIDSR/gcgdata/geneseq/geneseq/AA1991.DAT: *
13: /SIDSR/gcgdata/geneseq/geneseq/AA1992.DAT: *
14: /SIDSR/gcgdata/geneseq/geneseq/AA1993.DAT: *
15: /SIDSR/gcgdata/geneseq/geneseq/AA1994.DAT: *
16: /SIDSR/gcgdata/geneseq/geneseq/AA1995.DAT: *
17: /SIDSR/gcgdata/geneseq/geneseq/AA1996.DAT: *
18: /SIDSR/gcgdata/geneseq/geneseq/AA1997.DAT: *
19: /SIDSR/gcgdata/geneseq/geneseq/AA1998.DAT: *
20: /SIDSR/gcgdata/geneseq/geneseq/AA1999.DAT: *
21: /SIDSR/gcgdata/geneseq/geneseq/AA2000.DAT: *
22: /SIDSR/gcgdata/geneseq/geneseq/AA2001.DAT: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4225	100.0	784	19	AAW53570
2	4225	100.0	784	20	AA17417
3	2996.5	70.9	777	20	AAV32074
4	2860.5	67.7	783	20	AAV32073
5	2854.5	67.6	781	19	AAW57887
6	2854.5	67.6	781	20	AAV30143
7	2854.5	67.6	781	22	AA849400
8	2852	67.5	758	21	AA170978
9	2834	67.1	780	22	AA898659
10	2767	65.5	799	19	AAW57886
11	2767	65.5	799	20	AAV30142

12	2529	59.9	587	19	AAW57888
13	2529	59.9	587	20	AAV30144
14	2364	56.0	572	20	AAV32075
15	1490	35.3	751	21	AAV70980
16	1470	34.8	763	21	AAV70977
17	1432.5	33.9	756	21	AAV70975
18	1398.5	33.1	841	21	AAV70981
19	1394	33.0	750	20	AAV17418
20	1384	32.8	770	21	AAV70976
21	940	22.2	508	21	AAV70974
22	873	20.7	265	20	AAV32072
23	788	18.7	271	19	AAW57889
24	788	18.7	271	20	AAV30145
25	389.5	9.2	229	21	AA628629
26	263.5	6.2	193	21	AA628629
27	181	4.3	30	19	AAW53567
28	181	4.3	30	20	AAV17422
29	177.5	4.2	120	21	AA628630
30	111	2.6	1174	13	AA628635
31	109	2.6	1390	18	AAW21636
32	107.5	2.5	1584	20	AAV33727
33	107	2.5	487	19	AAW64182
34	107	2.5	488	19	AAW64169
35	107	2.5	1174	11	AA608257
36	107	2.5	1174	17	AA608257
37	107	2.5	1174	17	AA608257
38	106.5	2.5	905	22	AAW09043
39	106	2.5	1080	22	AAW09043
40	106	2.5	1637	22	AAW09043
41	105	2.5	19	19	AAW53568
42	105	2.5	19	20	AAV17423
43	104	2.5	933	22	AA664151
44	102	2.4	679	21	AA18224
45	101.5	2.4	478	20	AAW8533

ALIGNMENTS

RESULT 1	
AAW53570	standard; Protein: 784 AA.
AAW53570:	
AC	06-JUL-1998 (first entry)
DT	Cucurbit raffinose synthase.
XX	Cucurbit raffinose synthase.
DE	Cucurbit raffinose synthase.
XX	Cucurbit raffinose synthase.
KW	Cucurbit raffinose synthase; sucrose; galactinol.
OS	Cucurbit raffinose synthase.
XX	
PN	JP10084973-A.
PD	07-APR-1998.
XX	
PF	28-APR-1997; 97JP-0111124.
XX	
PR	26-JUL-1996; 96JP-0198079.
XX	26-APR-1996; 96JP-0107682.
PA	(AJIN) AJINOMOTO KK.
XX	
XX	WPI; 1998-264858/24.
DR	N-PSDB; AAV22250.
XX	
PT	Raffinose synthase gene - useful for preparation of raffinose in
XX	transformed plant
PS	Claim 3; Pages 17-20; 26pp; Japanese.
XX	
CC	The present sequence is cucumber raffinose synthase, which

Japanese artichoke
Amino acid sequenc
Rape seed raffinose
Wheat raffinose sy
Rice raffinose syn
Corn raffinose syn
Wheat raffinose sy
Soybean raffinose
Rice raffinose syn
Corn raffinose syn
Soybean raffinose
Corn raffinose syn
Amino acid sequenc
Arabidopsis thalia
Arabidopsis thalia
Cucurbit raffinose
Cucurbit raffinose
Novel toxin expres
Grapevine leafroll
Photographus lunin
Lettuce resistance
Lettuce resistance
B. thuringiensis to
B. l. toxin 81A2.
Bacillus thuringie
Rat CARD-6 protein
Human bone marrow
Human bone marrow
Cucurbit raffinose
Cucurbit raffinose
Arabidopsis thalia
Plasmodium falcipa
Salmonella enteric

CC forms raffinose from sucrose and galactinol, has an optimum pH of 6 to 8 and working temperature of 35 to 40 degrees C, has a molecular weight of 75 to 95 kDa by gel filtration or 90 to 100 kDa by PAGE and SDS-PAGE under reductive conditions and is inhibited by iodoacetamide, N-ethylmaleimide and myoinositol.

XX Sequence 784 AA:

Query Match 100.0%; Score 4225; DB 19; Length 784;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 784; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPSFRKNGSNVVSFDGLNDMSPPFAIDGSDFTVNGHSFLSDVPENIVASPSPTSIDKS 60
DB 1 mapsfkngsnvvsfdglndmsppfaidsdftvnghsflsdvpentivaspsptsidks 60
QY 61 PVSVCIVGFPDASEPDSRHVYSIGKLDIRFMSIFRFVWMTTHVWGRNGDLESETQIV 120
DB 61 pvsvcivgfdaasepdsrthvysigklkdirfmsifrfkvwmtthvwgrngdlesetqiv 120
QY 121 ILEKSDGRPVYFLLPIVEGPFRTSIQPGDDFVDVCESGSSKVVDAEFRSMLYLHAGD 180
DB 121 ileksdgrpyvflfpivegpfirtsipgdddfvdvcessgsskvvdasfrsmlylhagd 180
QY 181 DPFAIVKEMAKIVRTHLGTFRLLLEKTPPGIYDKRGWCTWDAFYLTVAHPGVIEGRHLV 240
DB 181 dpfalvkeamkivrthlgtfrllleektppgiydkrgwctwdaftyltvhpqviegvrhlv 240
QY 241 DGGCPGVLVIDDGWOSIGHSDPRTKGNNOYVAGBDMPCRLKFOENYKFRDYNPKA 300
DB 241 dggcpgvlvlddgwosighsdprtkgngnoyvagbdmprclkfoenykrdyvpka 300
QY 241 dggcpgpvlvlddgwqsighsdprltkegmngtyvageqmpcrltkfgenykrfdynpka 300
QY 301 TGPRAGOKGMAFIDELKGEFTEVHHVYVWHALCGYWGSLRPQVGLPEARVIOPLSPG 360
DB 301 tgpraqgkmafidelkgeftevhvyvwhalcgylwgsllrpqvglppearvioplspg 360
QY 361 LQMTMEDLAVDKIVLHKVGLVPRPEKAEMEYEGLAHLEKVGIDGVKIDVILHLEMLCEDY 420
DB 361 lqmtmedlavdkivlhkvglvprpekaemyeglahlekvgidgvkidyvilhlemledy 420
QY 421 GGRVDLAKAYTKAMTKSINKHFKGNGVYASMEHCNDFMFLGTEALSLGRVGDDEWCTDPS 480
DB 421 ggrvdlakaytkamtksinkhfkngvyaasmehcndfmflgtealislgrvgddewctdps 480
QY 481 GDPNCTFPLQCHWVHCANDSLMNGNFTHPDMFOSTHPCAFHAAARAIISGPIYSD 540
DB 481 gdpnctfplqchwhvncandslmngnfthpdmfosthpcafhaaaraisgpiysd 540
QY 541 SVGKHFDLKLKLVLPDGSILRSEYVALPTRDCLPEDPLHNGETMLKIMNLKFTGVIGA 600
DB 541 svghkfdlklklvlpdgsilrseyvalptrdclpedplhngetmlkwnlnkftgvlga 600
QY 601 FNCGGGMCRETRNOCQSQTSKRVTSKTNPKDITWHSGENISIEGVKTFALYLYQAKK 660
DB 601 fncgggmcetretrnocsqtskrvtsktnpkditiwshsgenisiegvktyfalyllyqakk 660
QY 661 LILSPDSODIDIALPPEELITVSPVTKLIQTSLHFAPIGVNMLNLSGATOSVDYDD 720
DB 661 lilspsodidialppeelitvsvptklitqtslhfapigvnmnlslsgatosvdydd 720
QY 721 LSSVEITGVKGCSEMRVFAKKPRACRIDGEDVGFYDDOMTVVOVPMFIDSSSGISYI 780
DB 721 lssveitgvkgscemrvfakpracridgedvgfyddomtvvovpmfidsssgisyl 780
QY 781 EYLF 784
DB 781 eylf 784

RESULT 2
AAV17417 standard; Protein: 784 AA.

XX AAV17417;
XX 29-JUL-1999 (first entry)
XX Cucurbit raffinose synthase.
XX Cucurbit raffinose synthase.
XX Raffinose synthase; sucrose; galactinol.
XX Cucumis sativus.
XX JP11123080-A.
XX 11-MAY-1999.
XX 24-OCT-1997; 97JP-0292969.
XX 24-OCT-1997; 97JP-0292969.
XX (AJIN) AJINOMOTO KK.
XX WPI: 1999-340516/29.
XX DR N-PSDB: AAX61238.
XX New raffinose synthase gene - for production of raffinose from
PT sucrose and galactinol
PS Claim 2; Page 25-27; 37pp; Japanese.
XX

CC The present invention describes a raffinose synthase, having an activity
CC of forming raffinose from sucrose and galactinol. The raffinose synthase
CC gene can be used for expression in a plant for the production of
CC raffinose. The raffinose synthase can give raffinose from sucrose
CC and galactinol efficiently. The present sequence represents raffinose
CC synthase from cucumber.
CC
XX
SQ Sequence 784 AA;

Query Match 100.0%; Score 4225; DB 20; Length 784;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 784; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPSFRKNGSNVVSFDGLNDMSPPFAIDGSDFTVNGHSFLSDVPENIVASPSPTSIDKS 60
DB 1 mapsfkngsnvvsfdglndmsppfaidsdftvnghsflsdvpentivaspsptsidks 60
QY 61 PVSVCIVGFPDASEPDSRHVYSIGKLDIRFMSIFRFVWMTTHVWGRNGDLESETQIV 120
DB 61 pvsvcivgfdaasepdsrthvysigklkdirfmsifrfkvwmtthvwgrngdlesetqiv 120
QY 121 ILEKSDGRPVYFLLPIVEGPFRTSIQPGDDFVDVCESGSSKVVDAEFRSMLYLHAGD 180
DB 121 ileksdgrpyvflfpivegpfirtsipgdddfvdvcessgsskvvdasfrsmlylhagd 180
QY 181 DPFAIVKEMAKIVRTHLGTFRLLLEKTPPGIYDKRGWCTWDAFYLTVAHPGVIEGRHLV 240
DB 181 dpfalvkeamkivrthlgtfrllleektppgiydkrgwctwdaftyltvhpqviegvrhlv 240
QY 241 DGGCPGVLVIDDGWOSIGHSDPRTKGNNOYVAGBDMPCRLKFOENYKFRDYNPKA 300
DB 241 dggcpgvlvlddgwosighsdprtkgngnoyvagbdmprclkfoenykrdyvpka 300
QY 241 dggcpgpvlvlddgwqsighsdprltkegmngtyvageqmpcrltkfgenykrfdynpka 300
QY 301 TGPRAGOKGMAFIDELKGEFTEVHHVYVWHALCGYWGSLRPQVGLPEARVIOPLSPG 360
DB 301 tgpraqgkmafidelkgeftevhvyvwhalcgylwgsllrpqvglppearvioplspg 360
QY 361 LQMTMEDLAVDKIVLHKVGLVPRPEKAEMEYEGLAHLEKVGIDGVKIDVILHLEMLCEDY 420
DB 361 lqmtmedlavdkivlhkvglvprpekaemyeglahlekvgidgvkidyvilhlemledy 420
QY 421 GGRVDLAKAYTKAMTKSINKHFKGNGVYASMEHCNDFMFLGTEALSLGRVGDDEWCTDPS 480
DB 421 ggrvdlakaytkamtksinkhfkngvyaasmehcndfmflgtealislgrvgddewctdps 480

DB 421 ggrvdlakaykamtkslnhfkngvlasamehndfmflgteaislgrvddfwctdps 480
QY 481 GDNRTFMLOGCHMVCANDSLWMGNFHPDMDFOSTHPCAFAFHAASRAISGPIYSD 540
DB 481 gdnrgtfvlgcmvncandslwmgnfhpdmdfosthpcafaaasraaisgpiyvsd 540
QY 541 SVGNHNFLLKLVLPDGSILRSEYVALPTKDLFEDPLHNGETMLKIMNLKFTGVIGA 600
DB 541 svgnhfnfllkvlpldgsilrseyvalptkdlfedplhngetmlkilmnlkftgviga 600
QY 601 FNGCGGCMRETRRNOCSGYSKRVTSKTRPKDIEHSGENPISIGVKFALYLOAKK 660
DB 601 fngcgggcmretrrnocsfgyakrvtsktrpkdiwhsgenpisiogvktfalylygakk 660
QY 661 LILSRPSODLIALDPFEFLITVSPVTKLIOTSLHAPIGLVNMLNTSGAIOISVDYDD 720
DB 661 lilsrpsodlialdpfeelivspvtkliotslhapiglvnmlntsgaioisvdydd 720
QY 721 LSSVEIGVKCGEMRVFASKRPRACRIDGEGVGFKYDQDMVVVQVWPIDSSSGISYI 780
DB 721 lssveigvkcgemrvfaskrpracridgedvgfkydqdmvvvqvwpidsssgisyl 780
QY 781 EYLF 784
DB 781 eylf 784
RESULT 3
AA132074
ID AA132074 standard; Protein: 777 AA.
AC AA132074;
XX
XX 17-JAN-2000 (first entry)
DE Mustard raffinose synthase.
XX
XX Raffinose synthase; mustard; transgenic plant.
OS Brassica juncea.
FH Key Location/Qualifiers
FT Misc-difference 210
ET /note= "encoded by ACR"
XX
XX EP953643-A2.
XX
XX 03-NOV-1999.
XX
XX 27-APR-1999; 99EP-0107430.
XX
XX 30-APR-1998; 98JP-0120550.
XX 30-APR-1998; 98JP-0120551.
XX 04-DEC-1998; 98JP-0345590.
XX 10-DEC-1998; 98JP-0351246.
XX
XX (SUMO) SUMITOMO CHEM CO LTD.
XX
XX Watanabe E, Oeda K;
XX
XX WPI: 1999-593144/51.
XX N-PSDB; AA220209.
XX
XX New sense and antisense genes, useful for altering the level of
XX raffinose in food plants -
XX
XX
XX Claim 26: Page 29-31; 55pp; English.
XX
XX This sequence represents mustard raffinose synthase, a protein
XX that can bind a D-galactosyl group through an alpha(1-6) bond to the
XX hydroxy group attached to the carbon atom at the 6-position of the
XX D-glucose residue in a sucrose molecule to form raffinose. cDNA
XX (see AA220209) encoding the enzyme was isolated from mustard

CC (Brassica juncea) leaf cDNA by PCR. Probes or primers generated from
CC plant raffinose synthase genes (see AA220207-10) may be used to obtain
CC other raffinose synthase genes by labeled detection or amplification
CC (claimed). These genes may be used to control the levels of
CC raffinose produced in plants. Antisense genes can be used to knock
CC out existing gene activity, and sense genes to increase the level
CC of gene activity. The resulting transgenic plants may be used as a
CC food source to alter the growing conditions for gut enterobacteria,
CC providing general health advantages.
XX

SO Sequence 777 AA:

Query Match 70.9%; Score 2996.5; DB 20; Length 777;
Best local Similarity 70.0%; Pred. No. 1.2e-274;

Matches 555; Conservative 99; Mismatches 114; Indels 25; Gaps 10;

QY 1 MAFSFKNGSNVVSFD---GLNDMSSP-FAIDGSDFTVNGHSFLSDVPENIVASPSY- 54
DB 1 map-----psvlsksdaavngidlsqkplfrlegsdlanghvvltcdvpnvvetaspyl 54
QY 55 TSIDKSPV--SVGCFVGFDA-SEPDNRHVYSICKLDIRMSTFRKRWMTTHWGRNG 111
DB 55 adkgepvdasagsifgindgeprsrhvaslgkldirfmslfrfkvwvltchwsgks 114
QY 112 DLSETOIVLEKSDSGRPVYVFLPIVEGPFRTSIOGDDPFVDCVSESSEKSVVDA5FR 171
DB 115 dlenecqllilenssgrfpyvlllpillegstfssfgqgeddvaavcsygcvtgsefr 174
QY 172 SMLYHAGDDPFALVKEAMKIVRTHLGTFRLEKTRPGIVDRCGCTWPAFLYVHPOG 231
DB 175 qvynvnaagddpfkvlvdamvvrvmhmtfklleektrpgivdrcgctwaflylvnpdg 234
QY 232 VIEGVRHLVDGCGPRGLVILIDGCMQSTIGHSDPRTYEGMNTYAGCEMPCRLKFOENYK 291
DB 235 vhgkvkclvdgscgcpjylvliddqwsghdsdgidvegmactvageqmpcrlkfgenf 294
QY 292 FRDYVNPKATGPRAGOGKMAFIDELKGEFTVEHHVYVHMLCGYWGRLPOYGPPEAR 351
DB 295 frdyvnpk----dknevgmfafrvdlkeefsvdyvlyvvhalscyggylrpgapclppst 350
QY 352 VIOPVSPGLQMTMEDLAVDKIVLHKVGLVPRPEKAEEMYGSLHAHLEKVIDEKIDVH 411
DB 351 ivrpelsgpjkltcmqdaavkivdtqigfvsppmanefyegllshlqnygidgvkvdvh 410
QY 412 LLEMICEDYGRVLDLAKAYYKAMTKSINKHFKNGVNASMEHCNDEMFLCTEASISLGRG 471
DB 411 llemlcekyggrvldlakaykaltssvnhfdgnyvasamehndfmflgteaislgrvg 470
QY 472 DDFWCTDPSGDPNGTFMLOGCHMVCANDSLWMGNFHPDMDFOSTHPCAFAHAAHSAI 531
DB 471 ddfwctdpsgdngtywlgcmvncaynaslwmgnfhpdmdfosthpcafaaasraai 530
QY 532 SGGPITYSDSVGNHNFLLKLVLPDGSILRSEYVALPTKDLFEDPLHNGETMLKIMNL 591
DB 531 sggpitysdsvgnhfnfllkvlpldgsilrseyvalptkdlfedplhngetmlkilmnl 590
QY 592 NKFTGVIGAFNCOGGGCMRETRRNOCSGYSKRVTSKTRPKDIEHSGENPISIEGYKTP 651
DB 591 nkytgyligafnccgggcmretrrnocsfgyakrvtsktrpkdiwhsgenpisiiegktp 650
QY 652 ALYLYOAKKILSKPSODLIALDPFEFLITVSPVTKLIOTSLHAPIGLVNMLNTSGA 711
DB 651 alflsqsckkvlisgpnndletlepfellivspvltlsgssvqfapglvmlntsga 710
QY 712 IOSVDDDDLSVEIGVKCGEMRVFASKRPRACRIDGEGVGFKYDQDMVVVQVWPID 771
DB 711 ioslvydhhe--sveigvkgemrvfaskrpracridgedvgfkydqdmvvvqvwpid 770
QY 772 SSSGGSIVIEYLF 784
DB 765 sapegslsskyef 777

Query Match	Best Local Similarity	Matches	Score	DB	Length
67.7%	66.6%	109	2860.5	20	783
109	109	132	9.7e-262	Indels	25
132	132	132	132	Gaps	12

QY	116	ETQAVILEKSDS--GRPVYELLPIVEGPRFTSIQPGD--DFVUVYCESSGSKVYDASFRS	172
Db	111	etqalllidsdesqgrpryilvllpiegpfiraslqpsvddvylcvesgsklvvgsitra	170
QY	173	MLYTHAGDDPALVKEAMKIVRTHLGTFRLLKEKTPRGVYDKRGWCTWPAFYLTVPQCV	232
Db	171	vlytragprdfklllktmckewqallgtfkllddctprgldvktgvcwdaftykvepyv	230
QY	233	IEGRHVLVDGCGPRGLVLIDGWMOSIGHSDPIT--KEGNOYTAGDOMPCRLKFOENK	291
Db	231	wegvkgylveugvprgylvlldgwsichddprrltdqeglnrtsaqgmprllkyeefk	290
QY	292	FRDVPVNPATKPR--AGCGMKAFIDELKGEKTYEHVYTHNALCGTWGILRPVPLPER	350
Db	291	frdykspriimghedhpnmgratfvyrdlkeefkvehyvvhatfgywgvprvprlpee	350
QY	351	RYIOPVLSPLQMTMEDLAVDKIYLHKVGIYVPRPEAKEMTEGSAHLHEKVGIDGVLDVI	410
Db	351	qvvefprklsprgilemmedlavdkivnnqiglvqpkageIyeglnshlencyiogvkvdyI	410
QY	411	HLLEMLCDYGGRYDLAKAYKAMTKSINKHFKNGVYIASMEHONDPMFLGTETASIGRV	470
Db	411	hllemmaedyggryvelakttykalttesvyrkhfkngvylasmeqndtmlygtetlclgry	470
QY	471	GDDEWCTDPSGDPNGTEWFLQCGHHVHCANDSLMWGNFIPHDMDFOSTHPCAAFHAASRA	530
Db	471	gddfprtpdpsgdngtyvlgqchmhvcaynalwgnfllhpddwdfgslnhpcaeefhaasra	530
QY	531	ISGGEIYVSDSVGKHNFDDLKVLVPGSILRSRYVALPTRDCLFEDPLANGFTMLKTN	590
Db	531	lsaggilysvdvvgkhnlpikrlkrladgsllrceyhalptkdcflvqplhdnglumlktn	590
QY	591	LNKFTGVIGANCOGGGCKRETRRNOCSOYKSVTSTKPNKDIEMGENPISIEGVT	650
Db	591	lnkfygvigvlnhcgggvstrsrxnmlctseykpslcktpskdvewgnhkprfkpyvec	650
QY	651	FALYVYQAKKILSKPSODLIDALDPEFEFLITVSPYTKLIQTSLHFAPIGLVNMLNTSG	710
Db	711	alQSVYDDDLSS--VELGVKGGCEMRVFAKKTRRACHIDGEDVGFKTDQDQVYQVP	768
QY	711	avksidisednedknvygikgagmmvyssekpkacrvcngedmeIey--eesmllkvgtw	768
QY	769	PIDSSSGGSIATYETLF 784	
Db	770	--nhnsggftfveyllf 783	
RESULT 5			
AAW57887	ID AAW57887 standard; Protein: 781 AA.		
XX	AAW57887;		
XX	AC		
XX	DT 23-SEP-1998 (first entry)		
XX	DE	Soybean raffinose synthetase.	
XX	KW	Raffinose synthetase; metabolism modification; food additive;	
XX	KM	gastrointestinal flora; soybean.	
XX	OS	Glycine max.	
XX	PN	EP849359-A2.	
XX	PD	24-JUN-1998.	
XX	PF	18-DEC-1997; 97EP-0122417.	
XX	PR	18-DEC-1996; 96GP-0338673.	
XX			

PA (SUMO) SUMITOMO CHEM CO LTD.

XX Oeda K, Mantanabe E;

XX WPI: 1998-324670/29.

DR N-PSDB: AAIV40801.

XX New nucleic acid molecule encoding plant raffinose synthetase -
PT capable of producing raffinose, used as food additives with
PT beneficial effects on gastrointestinal flora

XX Claim 1; Page 31-34; 44pp; English.

XX This sequence represents the soybean raffinose synthetase of the
CC invention. The raffinose synthetase is capable of producing raffinose by
CC combining a D-galactosyl group through an alpha (1-6) bond with a
CC hydroxyl group attached to the carbon atom at position 6 of a D-glucose
CC residue in a sucrose molecule. The DNA can be used to modify metabolism
CC of a host organism by introducing into the host organism or its cell so
CC that the content of the raffinose family oligosaccharides in the host
CC organism or cell is changed. Raffinose oligosaccharides are useful as
CC food additives with beneficial effects on the gastrointestinal flora.

XX Sequence 781 AA:

Query Match 67.6%; Score 2854.5; DB 19; Length 781;
Best Local Similarity 65.9%; Pred. No. 3.6e-261;
Matches 525; Conservative 123; Mismatches 120; Indels 29; Gaps 11;

QY 1 MAPSFKNGSGNVNVSFD--GLNDKSSPFAI--DGSDFTVNGHSLSDPENIVASPSPTYS 56
DB 1 mpsst-----skveinsfglvngnlpstlsgsnflanghpltevenpiltvpsp--- 53
QY 57 ID-KSPVS-----VCCFYGFDASEPDSSRHVVSIGLKIDIRKFSIRFYVWTTTHVGNR 109
DB 54 ldksskneddvdvvcfgvfghadeprsrhvaslgrklrgkifmslfrfkwvwlthvgsn 113
QY 110 GGDLESETQIVILEKSDS-GRPYVFLPPIVEGFRFRTSIQPGDDFVDCVSSGSKRVDA 168
DB 114 ghelehetqmmlldkndqigrpvlilpillaqasfiaslpglddydvcmesgstrvcgs 173
QY 169 SFRSMYLTLAGDDPFALVTEAMKIVRTHLGFRLLEKTPPGIVDFKFGCTMDAFYLVH 228
DB 174 sfsgclylvnvhndpygltreatkvvrmhlgftfkleeekapvldkfgyctwdaflxvh 223
QY 229 PGQVIGVNRHLVDGCGPGLVLIDGMQSIGHSDPIT-KEGNQTVAGEQMCRLLRQ 287
DB 234 psgvwegvkvlgvegscppmvlllddgwgalchdegrltdqegmkrtcsaqemprcrlvk 293
QY 288 ENKFRDYVNPRTGFRACOKGMAFIDELKGFRTVENHYVNAALCGYWGCLRPOVPL 347
DB 294 enkfifqysgkds-----ekgmgaftvrdlkqgfrsvegyvwhalcgyvgvrvpvgm 348
QY 348 PEARVIOPLVPELOMTMEDIAVDKILVKHVGSLVPEKAEEMVGAHLHEKGIQVVKI 407
DB 349 pqekvvtprklsnglkltmkdlaavdklvsngvgvlprhlalhllgyhlsrlesagldgvkv 408
QY 408 DVTHLLEMLCEDYGGRAVDLAKAYKAMTKSINKHFKNGVGIASMEHCNDFMLGTEAIST 467
DB 409 dvthllemlseeygrvteakaykaltasvkkhfkngvliasemehcndfllgteaial 468
QY 468 GRVGDFWCTDPSGDPNGTFWLOGCHMHNICANDSLMKNFNIHDMQFQSTNHCALFHAA 527
DB 469 gnygdftwctdpsgdpngtfwlogchmhnvcaaynsilwmgfllqpdwcmfsgthpcaefhaa 528
QY 528 SRAISGPIYVSQVSKHNPDLKLVLPOGSLRSEYVALPRLDCLFEDPRLNGEFTMLK 567
DB 529 sraissgpiyvsqvkhnpldkllksalprgcllrcqhyalprtdclfeepldngctmlk 568
QY 588 IWNINFTGVIGAFNCGGGMCRRETRNOCFSQYSKRVTSKTPKDIEMHSGENPSISIEG 647
DB 589 lwnlnftgvigafnncgggmcrretrnncfsqyskrvtsktpkdienmhsgenpsisic 648

QY 648 VKTFALYIQAOKKLITSRKPODLIALDPFEFELITVSPVTKIQTSLHAFIQLVNLN 707
DB 649 mnvtaaylftkdhklklmkasekleveleptfeiltvsvlylskklldfaplglymnl 708
QY 708 TSGAIOVDYDDLSVEIGVAGCGEMRVFASKRPACRIDEDVGFKKYDDQMVVVOVP 767
DB 709 tsgaiovdedyddlsveigvagcgemrvfaskrppacridedvgyfkkdyddqmvvvovp 767
QY 768 WPIIDSSGCIIVIEYLF 784
DB 768 wp---saaskismveflf 781

RESULT 6
AAV30143
ID AAV30143 standard; Protein: 781 AA.

AAV30143;

26-OCT-1999 (first entry)

Amino acid sequence of a raffinose synthase protein.

Raffinose synthase; plant; sucrose; raffinose.

Glycine max.

JP11215984-A.

10-AUG-1999.

12-DEC-1997; 97JP-0342899.

28-NOV-1997; 97JP-0329006.

18-DEC-1996; 96JP-0338673.

(SUMO) SUMITOMO CHEM CO LTD.

WPI: 1999-511112/43.

N-PSDB: AAZ10002.

New raffinose synthase gene - is prepared from a plant material

Claim 8; Page 25-27; 40pp; Japanese.

The present sequence represents a raffinose synthase protein. The
CC sequence is isolated from plant material. The protein forms raffinose
CC by complexing alpha(1 to 6)- D-galactosyl hydroxyl group of the 6C
CC of D-glucose residue in sucrose molecules.

Sequence 781 AA:

Query Match 67.6%; Score 2854.5; DB 20; Length 781;
Best Local Similarity 65.9%; Pred. No. 3.6e-261;
Matches 525; Conservative 123; Mismatches 120; Indels 29; Gaps 11;

QY 1 MAPSFKNGSGNVNVSFD--GLNDKSSPFAI--DGSDFTVNGHSLSDPENIVASPSPTYS 56
DB 1 mpsst-----skveinsfglvngnlpstlsgsnflanghpltevenpiltvpsp--- 53
QY 57 ID-KSPVS-----VCCFYGFDASEPDSSRHVVSIGLKIDIRKFSIRFYVWTTTHVGNR 109
DB 54 ldksskneddvdvvcfgvfghadeprsrhvaslgrklrgkifmslfrfkwvwlthvgsn 113
QY 110 GGDLESETQIVILEKSDS-GRPYVFLPPIVEGFRFRTSIQPGDDFVDCVSSGSKRVDA 168
DB 114 ghelehetqmmlldkndqigrpvlilpillaqasfiaslpglddydvcmesgstrvcgs 173
QY 169 SFRSMYLTLAGDDPFALVTEAMKIVRTHLGFRLLEKTPPGIVDFKFGCTMDAFYLVH 228
DB 174 sfsgclylvnvhndpygltreatkvvrmhlgftfkleeekapvldkfgyctwdaflxvh 223

QY	229	PGCVIEGVNRHLVDGCGPREGVLIDGWSGISTGHSDDPT-KEGMNGTYAGGOMRCRLIKXQ	26
Db	234	psgvvexvqglveggcpbmvlldqwgqalchdepllddegmkrtseagmpcorlvkkl	29
QY	288	ENKFPDYNNPKATGPRBAQCKGMKAFIDELGEFKEVHEVYVWHALCGYWGGLRPVPGH	34
Db	294	emkrtfrygsgskds-----ekmgafivdlkeqftriseeqyvvwihalcgysgyvtrpkrpvm	34
QY	348	PEARVIOPLVSPGLQMTIMDLADVRLHKVGLVPRPEKAEMEYEGHLAEKVGIDGKVI	40
Db	349	pgakvvtprklsnglklitmkladvklvansvgvlvprhlahllyegllhsrtlesagldgvkv	40
QY	408	DVTHLEMLCEDYGGGVNDLAKAYYKAMTKSINKHKNGCVIASMEHGNDDMPFGTEAISL	46
Db	409	dvthlllemlseeyggrvelakaykallasyvkkhkngvliasmehcndflllgteaisl	46
QY	468	GRGDDPFWCTDPFGSDNGNGFWMLOGCHMYHCANDSLMNGNFTHPPMDMFQSTHPCAFHAA	52
Db	469	grvgddftwctcdpsgdqngcyvlgqchmwhcyanslmngnffgqwdmfgsthpcaeethaa	52
QY	528	SRAISGGPIVSDSVGKHNFDLKLVLDPDGSILRSEYALPTROCLFEDPLANGETMLK	55
Db	529	sraisgppvysdcvqkhnfkllkslalpdqfcllrcqhyalptrdcclfedplnhqktmlk	55
QY	588	IWNLNKRTGYIGAFNCGGSGWGBRETRBNCCFQYSKRYTSTKNPKDLEMHSGENPISIEG	64
Db	589	lwnlnkrylgllfncqgggwcpcvtrtnksaselsqvltvclaspqdlsewngskspclcig	64
QY	648	VKTFEATLVYAKKLLILSKPSODIALDPFEFELITVSPVTKLIOTSILRAPIGLVNMLN	70
Db	649	mvrfayvllfdhklklmkaseklevslpftfclltvsvpvlvlsklllgfapiglvmnl	70
QY	708	TSGAIGSVYDDDLSSVEYLGWKGCGGMARVPASKRRARIRIDEGVGFXYDDQWVYVQVP	76
Db	709	tsgaigsmetdnhldvkvklygvsgemkvfasekpvscldgvyvvkldy-edkmlrvqvp	76
QY	768	WPIDSSGGISVIEYLF 784	
Db	768	wp----saskismwelf 781	
RESULT 7			
AA	ABA49400		
AC	ABA49400	standard; protein; 781 AA.	
XX	07-MAR-2001	(first entry)	
XX		Soybean raffinose synthase.	
XX		Plant promoter; transgenic plant; desired property.	
XX		Glycine max.	
XX	EP1048733-A2.		
XX	02-NOV-2000.		
XX	27-APR-2000;	2000EP-0108962.	
XX	30-APR-1999;	99JP-0124527.	
XX	01-SEP-1999;	99JP-0247211.	
XX	(SOMO) SUMITOMO CHEM CO LTD.		
XX	Ishige F, Watanabe E, Oeda K;		
XX	WPI: 2001-104537/12.		
XX	N-PSDB; AAC89523.		
XX	New soybean plant promoters useful for generating transgenic plants		

XX	with desired properties	-
PS	Example 6: Page 24-27; 36pp; English.	
XX	The present invention provides novel plant promoters which can be used in	
CC	the production of transgenic plants which express genes with desired	
CC	properties.	
XX		
XX	Sequence 781 AA:	
50		
	Query Match	67.6%; Score 2854.5; DB 22; Length 781;
	Best Local Similarity	65.9%; Pred. No. 3.6e-261;
	Matches 525; Conservative 123; Mismatches 120; Indels 29; Gaps 11.	
OY	1 MAPSKNGSNVVSFD--GLNDMSPPAI--DGSDFYVGHSLSDVPENIYASPPYTS	56
Db	1 mapei-----sktvelnsfglvngnlpstllesnflanghpflltevpennlvtvpsp---	53
OY	57 ID-KRPVS-----VGCVRPDSAPDSHHVYSIGLKDRIEMSIPIRFKVMWTTTHVGN	109
Db	54 idaksknneddavvgctvgfghadepstrshvaslglvrlkmslfrfkwvttwhvggn	113
OY	110 GDSESEMOIVLEKSDS-GRPYVFLPIVESEPFRTSIOPGDDDFYDVCVSGSSKRVDA	168
Db	114 ghelehegmmldkndqgrpfvllrlpgsfasldpjdldgydvcmesgstfvcgs	173
OY	169 SFRSMVLHAGDPPALVKEAMKIVRTHLIGTRLLLEKTRPGIVDVKFGCTMDAFLLVH	228
Db	174 sfqsclythvghdpyqllrtaekvnmthlygfklleektapvldkfvcwtwda fylkvh	233
OY	229 POGVIEGRVHLDVGGCRPELVLDGWSIGHDSPIIT-KEGNNOVAVAGHOMPRLKPO	287
Db	234 psgvewgyvlgiveggcprpmvllddwgalcndepildqegmkrtcsagegmprclvkle	293
OY	288 ENYKRPDYVNPKATPRPRAQOKMKAFIDELKGBFKTVEHVVYVHVALGYYGLRPVPG	347
Db	294 enykrtfrycsqkds-----ekgmgaivtrldkegfveevgyvwhalcgywgyvtrpvgm	348
OY	348 PEARVIGPVLSFGLQMTMEDLAVDKTLVLRKGLVPRPKAEEMTEGHLNLEKVGIDGVI	407
Db	349 pqakvvtprlslnglkrktmkldavdklvsngvdlvprhlahllyegllhstresagldgvkv	408
OY	408 DYIHLLMLCEGYGGVVDLAKAYYKAMKTSINKHFKFGNGVYIASEMHCNPFMFAGTASL	467
Db	409 dvihllemseevggvtelakayykalasvkkhfkngnyiasemeicndffllqtealal	468
OY	468 GAVGDDFMCDESGDPNGTFLWLGCHMVCANDSLWMGNFIHPDWMFOSTHPCAFAHAA	527
Db	469 gtyvgddfwctdpsgdngngyvlqgclmnhcaynslvmgnfllqgdwdfgsthpaetfha	528
OY	528 SRAISGPIYVSDSVGKHNFDLKIYLVDPDGLSLSEYYALPIPRDCLFEDPLINGEMTK	567
Db	529 sraissgpyvsvdcvqkhnfkllksialpdgltllcqhyalprdcifedplindgktmlk	568
OY	588 IYNLNFTEVIGAFNQGQGGWMCERTRNOCFSQYSKRVMSKTTPKDIEMWENSIIEG	647
Db	589 iynlnfyvrlglfncqgggwwcvttrtnsaaserisqvtclaspqdtewngsplicikg	648
OY	648 VTFPALYLOAKKLLIISKSSODDIALDPFEPLIIVSPYTKLIQTSLHAPIGLVMLN	707
Db	649 mvfayvlltkdhkiklmksekelsleptfcllcvspvialskrlifagpdlgvmln	708
OY	708 TSGAIVSDVDDDLSSVEIGVKCGEMRVFASKPRACRIDGEDVGFKYDQDMVYVQVP	767
Db	709 tsgaigsmefdnhdvklvgvsgcmkvfasekpsckldgyvvvklfy-edkmlrtvqv	767
OY	768 WPIDSSSGGIVTELYLF 784	
Db	768 wp-----sasklsmveflf 781	

AAV0978
ID AAV0978 standard: Protein; 758 AA.
XX
AC AAV0978;
XX
DT 09-AUG-2000 (first entry)
XX
DE Soybean raffinose synthase from clone sfl1.pk125.d4.
XX
KM Soybean; raffinose synthase; raffinose saccharide;
XX clone sfl1.pk125.d4; nutritional; soy protein.
XX
OS Glycine max.
XX
PM W0200024915-A2.
XX
PD 04-MAY-2000.
XX
PF 22-OCT-1999; 99WO-US24923.
XX
PR 23-OCT-1998; 98US-0105451.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E. I.
XX
PI Allen SM, Hltz WD;
XX
XX WPI: 2000-350754/30.
DR N-PSDB: AAD00335.
XX
XX
XX Nucleic acids and encoded proteins involved in the biosynthesis of
PT raffinose, useful for producing soybean seeds with a reduced raffinose
PT content and therefore improved nutritional quality -
XX
PS Claim 2; Page 47-49; 58pp; English.
XX
XX
XX The present sequence is a raffinose synthase from
CC clone sfl1.pk125.d4 isolated from a soybean immature flower cDNA
CC library sfl1. Raffinose synthase is involved in the biosynthesis
CC of raffinose and higher homologues in the raffinose saccharide family
CC from sucrose. The present sequence is useful for reducing the raffinose
CC saccharide content of soybean seeds which improves the nutritional
CC quality of the soy protein products derived from them.
XX
SQ Sequence 758 AA:

Query Match 67.5%; Score 2852; DB 21; Length 758;
Best Local Similarity 67.5%; Pred. No. 5,9e-261;
Matches 530; Conservative 94; Mismatches 133; Indels 28; Gaps 9;

QY 1 MAPSFRNGSGSNVSPFGDNDMSFPAIDGSDFTVNGHSPISDVPEINIVASPRYSIDKS 60
DB 1 mgspsakkaak-----sgvckhmkgfslcnslkvngqvllsqpkvcllpcctydt---- 52
QY 61 PVSVGCFVGFDPDASEPDSRHVYSIGKLDIRFMSIFRFKVMYTHWYGRNGDLESEFOIY 120
DB 53 -htgcflgfhnapksrtnvapjgqlnklsfcsiffrkwwtllwgsngrdletetqfl 111
QY 121 ILEKSDSGRPYFLLPIVEGPFRTSIQPGDDDFVDVCESSSKVVDASEFRMLYLHAGD 180
DB 112 mlg-----shpylflrlpqlpfraalqphaddnavaveesgashvclassfdvylhagd 167
QY 181 DPAALYKEMAKYRTHLGRFLLEKTPPGIYVRFGMCTADAYLVHPOGVLEGRHLY 240
DB 168 nptllykeamrvvtrahlgfklleeklvpgmwdkfwtctdaflylvhpegvregykglyv 227
QY 241 DGSGPGLVILIDPGMOSIGSDPIRKEGNNOTVAGEOMPICRLKFOENKFFDYNPKA 300
DB 228 dggcpdpgfvlldggwqclshdsdp-ekegmnglvageqmpcrlisyeenklrtsy----- 281
QY 301 TGPRAGOKGAKAIDELKGEFKTVEHVYVWHALCGYWGGLRPQVGLPEARVLIQVLSPG 360
DB 282 ---keg-kgjkgfvrclkeefgeveyvwhalcygyvgvyrpvgmaaeaaevckpkltg 337

QY 361 LQMTMEDLAVDKIVLHKVGLVPEKEAEMYEGILHAHLKEVIGIDGVKIDVYIHLEMLCEDY 420
DB 338 lqymedlavdkivmgvgvvppeylvgemyeqlhahlesagldgvvvdvylhlemckxy 397
QY 421 GGRVDLAKAYYKAMTKSINKHFKGNGVIAEMHCNDFMFLGTEALISLGRVGDDEFMCTDPS 480
DB 398 ggrvdmakayykalctasvvrkhfkngvlasmebchndfmllgteaalslgrvgddfwctdpy 457
QY 481 GDPNGTFWLQGCWVHCANDSLMGMNFHPDMDMFOSTHPCAAFHAAASAISGPIYVSD 540
DB 458 gdpngtfwlgcwmvncaynsllwmgnflhpdmfmgsthpcaafhaasaisgpiylisd 517
QY 541 SVGKHNFEDLLKRLVLPDGSILRSEYALPTRDLPEDPRLHNGFTMLKTNLKNKFTGIVA 600
DB 518 tvgnhmfellklclalpdgsilrcehyalprdccladpilhgcukmklvmlnvgvlgv 577
QY 601 FNOCGGMCRETRRNOCFQOYSKRVTSKTNPXKDIEHSGENPISIGVTFALYLYOARK 660
DB 578 fncgggwftrrlsrnkcaaeefsrvtcknlkldewsgknpslsgvqlfsayfsqak 637
QY 661 LILSKPSDL-DIADLPFEFELITVSPVTKLIOTSLHFAPIGLVNNLNTSGALQSVDYD 719
DB 638 lilspsdseelslepfnfelitvsvptvlpqsvkfapiglvmmlnvgvsgjafde 697
QY 720 DLSSVEIGVKGSGEMRVFASKKPRACRACRIDGEDVGFXYDDQDMMVYVQVPMIDSSSGISV 779
DB 698 gqnlvevgilrtgcmrvvasexprctridgkevdfey-egsmvnlqvpwv---gsasklsc 753
QY 780 IEXLF 784
DB 754 vqyvf 758

RESULT 9
AAB98659
ID AAB98659 standard: protein; 780 AA.
XX
AC AAB98659;
XX
DT 17-AUG-2001 (first entry)
XX
DE Soybean protein: SEQ ID 1.
XX
XX Mutant; mutein; raffinose synthase; raffinose oligosaccharide reduction;
KM Plant; soybean.
XX
OS Glycine max.
XX
PN JP2001078783-A.
XX
PD 27-MAR-2001.
XX
PF 03-JUL-2000; 2000JP-0200571.
XX
PR 09-JUL-1999; 99JP-0196036.
XX
PA (SUMO) SUMITOMO CHEM CO LTD.
XX
DR WPI: 2001-313373/33.
XX N-PSDB: AAH27438.
PT Novel mutant protein of raffinose synthase is useful for reducing the
PT raffinose oligosaccharide content in a plant body -
XX
PS Disclosure: Page 18-20; 30pp; Japanese.
XX
XX The present invention relates to a mutant protein of raffinose synthase
CC in which at least one aromatic amino acid present at the position of
CC about 1-7 amino acids from the N-terminus is deleted or replaced. The
CC mutant protein can be used for reducing the raffinose oligosaccharide
CC content in a plant body. The present protein from soybean, was used in
CC the present invention.

OY	411	HLMLCEDVCGRDVLKAAVKAMTKGSLNHPKNGVIASEHONDPMFTEAISIGRV	470
Dd	430	hllellseeyaggrtelareykaltsstsvkhhfyngvfaasmehncndflllgteaistigrv	489
OY	471	GDDFWCTDPGSGDPRGTFMALGSGCHHVHCANDSLMWMGNFIHPDMDFOSTHPCAAFHAASRA	530
Dd	490	gddctwscadpsgdprngtywlgscmhvhcaaylsmgnfigdpdkmfgsthpcaeftnastra	549
OY	531	ISGGPIYVSDSVGKHNFDDLKLVLPGSILRSEYVALPTROCLFEDPLNHGETMLKIWN	590
Dd	550	ls99pilyasdcvgnhnfhklkslyldpgslilrcqhyalprcdclfedplnmgtmkltwn	609
OY	591	LNRKTGYIGAANCOCGGGMCRTRNOCFSQYSKRVTSKTPMKOLEMHSGENPISIEBVRT	650
Dd	610	lnklytgyiglincc9gg9vcpearrtnkvsaelstravtcyaapedlewcnqgtlmstkyvdf	669
OY	651	FALYLVAOKAKLITLSKPSODDIALDPEEFELITYSPYTKLIQNSLTHAPRGLVMNLWTSG	710
Dd	670	fayvfefkekktlrlnkcscdrklvsllepstsfelmetyspvkfskriftdaprlgylvmlnsgs	729
OY	711	AIGSVYDDDLSSVEIGVKGCGEMERVFASKPRACRIDGEDGVGEKYDQDOVVVYWVPWPI	770
Dd	730	aigsltefdadaslvklivrgvgsenvfasekpvcckldgykvkfily-edkmarvtglwp-	787
OY	771	DSSSGGISIVETLF 784	
Dd	788	--ssaslsvglflf 799	

RESULT	11
AAV30142	
ID	AAV30142 standard; Protein; 799 AA.

AC	AAV30142;
XX	
DT	26-OCT-1999 (first entry)
XX	
DE	Amino acid sequence of a raffinose synthase protein.
XX	
KM	Raffinose synthase; plant; broad bean; sucrose; raffinose.
XX	
OS	Vicia faba.
XX	
PN	JPI1215984-A.
XX	
PD	10-AUG-1999.
XX	
PF	12-DEC-1997; 97JP-0342899.
XX	
PR	28-NOV-1997; 97JP-0329006.
XX	
PR	18-DEC-1996; 96JP-0338673.
XX	
PA	(SUMO) SUMITOMO CHEM CO LTD.
XX	
DR	WPI: 1999-511112/43.
DR	N-PSDB; AAZ10001.
XX	
PT	New raffinose synthase gene - is prepared from a plant material
XX	
PS	Claim 5; Page 19-21; 40pp; Japanese.
XX	
CC	The present sequence represents a raffinose synthase protein. The
CC	sequence is isolated from plant material of broad beans. The
CC	protein forms raffinose by complexing alpha(1 to 6)- D-galactosyl
XX	hydroxyl group of the 6C of D-glucose residue in sucrose molecules
XX	
XX	Sequence 799 AA;

Query Match	65.5%	Score 2767;	DB 20;	Length 799;
Best Local Similarity	64.0%;	Pred. No. 7.3e-253;		
Matches 508;	Conservative 118;	Mismatches 140;	Indels 28;	Gaps 8

```

0Y 11 NVSPSDGLNDKSSP-FPI---DSDSEFVNASHSLFDVPEKIVNASPSPIYSI-----D 58
Db 14 dvlstldtldgngnslfslctldqsrdflianghrltqpprltcttcttcaasflnlkskd 73
0Y 59 KSP-----VSVCFCFGFDASEPDSRHVVSIGLKXDIRMSJFRFXWMTTTHAVGRNG 111
Db 74 tlpnnntemllqgcfcfvgfnstpeckshvvrplgkllgkfmalsfrfkwvwtclhwgtng 133
0Y 112 DLESTQIVLEKSDS-GRPVYFLPLPIVEGPFRTSIOPGDDDFVDCVBSGSKVVDASE 170
Db 134 elqhetgmllldkxdsigrpyvlllpillettfsrlqplndhllgmvsaeashtvgsstf 193
0Y 171 RSMJYLAHAGDFFALYVEAMKIVTHTHGTSTRILEEKPFGIVYVKFGMCTMDATAYLVNHP 230
Db 194 kacjylhsmndpyllleavvnlvqtqlgtlctleekkapalldkfgvctdaaylklvhr 253
0Y 231 GVIEGRHLVDGCGPGLVYLIDDMQOSIGHDSPEITKEGKNQTVAGEBOMPCRLKFOEY 290
Db 254 gwvegvkeldtgdgpppflvlllddgwslchdddeddsnmrtsaegqmpcrltvkyeas 313
0Y 291 KFRDYVNPKATGPRAGOKMAKAFIDELKGEKTYENHYVNNALCGTWGLRQVPGLEPA 350
Db 314 kfireyempe---ngrkkgj19gfvrdrlkeefsgsvesyvvhalccywgavrgpvhgmprk 369
0Y 351 RVIOPLVSPGLQMTMEDLADYVKIYLHKVYGLVPPREKAEEMEGJLAHLEKYGIDGVIDY 410
Db 370 rrvvpkpsvsgjklmctmedavdklvengvgtlyppdfahemfdgllshlesagldgvkxvdy 429
0Y 411 HLEMLECEDYGGRYDLAKAYYKAMTKSINRHKFGNGVYASMEHCNDPMFLGTBAISLGRY 470
Db 430 hlllelseeygravelarraykaltssvkkhfkngytlasemchondffllgttaalslgy 489
0Y 471 GDDPACNDPSSDPRGCTFWLQSCCHVHCANDSLMMGNTIHPDMFOFSTHPCAFPHASRA 530
Db 490 gddfwcsdpdsgprngcylwlgqclmhvhaaynslmwngnldqdwlmfgeclhpcaefhaasra 549
0Y 531 ISGGPIYVSDSVGRHNNDLKLVLPGSILSRSEYVLPDRCLFEBPLRNGCTETMLKYN 590
Db 550 lsgpplysdcvgnhntfllkslslpdsallrcghylprtdclfeoplmgntmlkyn 609
0Y 591 LNKFTYGVTAFCNCGGCGKCRPTRNOCFSQYSKRVYSKTYNPKDIEWHSGENPISIECVKT 650
Db 610 lnxtyvgvlgllfncgvggwcpearrnkvsaeisfzavtclaaapedewngktcpmstkgvdf 669
0Y 651 FALYVYOKKKIILSKPSDDIALDRPEFELITYSPUTKLIQNSLHAPRGLNMLNTSG 710
Db 670 faayffkckkrlmkcsdrklvstlepfstfelmclvspkvfskrlfdqaplyovvmlns9 729
0Y 711 AIQSVYDDDLISVEIGKVGCGEEMRVEASKRPACRACRIDGEDVGEKKYDQDQVVVQYPMPI 770
Db 730 aigstlefadnaslwkjgyvgcgemsfvashekvpcckldgykvkfly-edkmaevqjllw- 787
0Y 771 DSSSGSIVIEYLF 784
Db 788 --ssstslvqflf 799

```

RESULT	12
AAW57888	
ID	AAW57888 standard; Protein; 587 AA

AC AAW57888;
XX
DT 23-SEP-1998 (first entry)
XX
DE Japanese artichoke raffinose synthetase.
XX
KM Raffinose synthetase; metabolism modification; food additive;
XX gastrointestinal flora; Japanese artichoke.
OS Stachys sieboldii.
XX
PN EP849359-A2.
XX

XX 24-JUN-1998.
PD
XX
XX 18-DEC-1997; 97EP-0122417.
PF
XX
XX 18-DEC-1996; 96JP-0338673.
PR
XX
PA (SUMITO) SUMITOMO CHEM CO LTD.
PI
XX Oeda K, Wantanabe E;
PI
XX WPI: 1998-324670/29.
DR
DR N-PSDB; AAV40802.
XX
XX New nucleic acid molecule encoding plant raffinose synthetase -
PT capable of producing raffinose, used as food additives with
PT beneficial effects on gastrointestinal flora
XX
XX Claim 1; Page 36-38; 44pp; English.
XX
XX This sequence is the Japanese artichoke raffinose synthetase of the
CC invention. The raffinose synthetase is capable of producing raffinose by
CC combining a D-galactosyl group through an alpha (1-6) bond with a
CC hydroxyl group attached to the carbon atom at position 6 of a D-glucose
CC residue in a sucrose molecule. The DNA can be used to modify metabolism
CC of a host organism by introducing into the host organism or its cell so
CC that the content of the raffinose family oligosaccharides in the host
CC organism or cell is changed. Raffinose oligosaccharides are useful as
CC food additives with beneficial effects on the gastrointestinal flora.
XX
XX Sequence 587 AA;
XQ

[illegible]

```
QY 649 KTFALYVQAKLLSKPSQDDIALDPFEELITVSPVKL 6300
      |||||: |||:|||| :|| :||:||||||| |
Db 540 ktfaYlfnekkvllskpskdiddtlpfdfeelitvsvkrl 5810
```

RESULT	13
AAV30144	
ID	AAV30144 standard; Protein; 587 AA

AC AAY30144;

DT 26-OCT-1999 (first entry)

DE Amino acid sequence of a raffinose synthase protein

KW Raffinose synthase; plant; sucrose; raffinose

OS *Stachys sieboldii*.

PN JP11215984-A

PD 10-AUG-1999.

PF 12-DEC-1997; 97JP-0342899.

PR 28-NOV-1997; 97JP-0329006.

XX

XX

DR N-PSDB; AAZ10003.

PT New raffinose synthase gene - is prepared from a plant material

PS Claim 12; Page 30-31; 40pp; Japanese.

CC The present sequence represents a raffinose synthase protein. The

CC by complexing alpha(1 to 6)- D-galactosyl hydroxyl group of the 6C

XX

Query Match	59.9%	Score 2529	DB 20	Length 587
Best Local Similarity	77.0%	Pred. No. 1.6e+230		
Matches 448	Conservative 62	Mismatches 70	Indels 2	Gaps 2

QY 109 NGDLESETQIVILLEKSDSGRPVFLPIVEGPERTSIQPGDDDFVDVCVESGSSKVDA 166

Db 2 ngsdleretqivldksdd-rpyivlplieqfraslqpvddfidi vesgstkves 60

QY 169 SFRSMY LHAGDDPFALVKEAMKIVRTHLGTFRLL E E K T P P G I V D K F G W C T W D A F Y L T V H 22

Db 61 straslymhagddpftlvkdavkvarhlgtrllleektppgividkfgwtwda fylnvq 1z

QY 229 PQGVIEGVRHLVDGCPGLVLIDDGWSIGHDSDPITKEGMNQTVAQEOMPCRLKFOE 28

Db 121 phgvmevgvlvdgscppglvllddgwsichndalltegmgrtsageqmpcrlktee 18

QY 289 NYKRDYVNPKATGPRAGQKGMAFIDELKGEFKTVEHVYVWHALCGYWGGLRPQVPLP 34

Db 181 nykfreyespnktgp-gpntgmgaFirdmknfksvdyvwhalcgylrpnvpglp 23

QY 349 EARVIQPVLSPLQMTMEDLAVDKIVLHKVGLVPPEKAEEMEYEGLHAHLEKVCIDGVKID 400

Db 240 eakliepkltpgikttmedlavdkivngvglvppfeveqmyeglhshlesvgldgvkvd 25

QY 409 VILHLEMLCEDYGGVLDLAKAYYKAMTKSINKHFKGNGVIA SMEHCNDFMFLGTEAISLG 400

Db 300 vihllemcedygrvdlakaykalsssvnhfnngvialglehcnclfmflyteaitlg 35

```

OY 469 RVGDDEFWCTDPGSDPNGTFTMLQGCNHYHCAANDSLMGMNFIRHPDWMFOSTHPCAFHAA5 528
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 360 rygdftwcddpsgdprgtfkwlgchmhcaynslmgnflhpdkwmfgsthpcaeethaa5 419
OY 529 RAISGPRIVSDSVGKHNNDLKLVLPGSILRSEYVLLPFRDCLFEDPLNGETMLKI 588
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 420 raiaggpilyvasvsgkhntellrslylpgsallrcdyalptrdcilfeoplhngktmlk1 479
OY 589 MNLNKFTGIGAFNCGGGMCNRETRNOCESQYSKRVTSKTPNKDIEHMGSENPISIEGV 648
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 480 wmynkftgyvgtfncgg9g9wrevrtngcaaeyslvassagpsdlemwqyspladvgv 539
OY 649 KTFEALYQAKKLILSKPSQDDIALDPFEFELITVSPYTKL 690
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 540 ktfaelylfnektivlekpsdkidltlepfdfeltvsvtkl1 581

RESULT 14
AAV32075
ID AAV32075 standard; Protein; 572 AA.
XX
AC AAV32075;
XX
DT 17-JAN-2000 (first entry)
XX
DE Rapeseed raffinose synthase.
XX
KM Raffinose synthase; rapeseed; transgenic plant.
XX
OS Brassica napus.
XX
FH Key Location/Qualifiers
FT Misc-difference 129 /note= "encoded by GGT"
FT Misc-difference 132 /note= "encoded by GGT"
FT Misc-difference 132 /note= "encoded by GGT"
FT Misc-difference 143 /note= "encoded by GGT"
FT Misc-difference 143 /note= "encoded by GGT"
FT Misc-difference 144 /note= "encoded by CCS"
FT Misc-difference 144 /note= "encoded by CCS"
FT Misc-difference 148 /note= "encoded by TCR"
FT Misc-difference 148 /note= "encoded by TCR"
FT Misc-difference 148 /note= "encoded by CGR"
XX
PN EP953643-A2.
XX
PD 03-NOV-1999.
XX
XX
XX 27-APR-1999; 99EP-0107430.
XX
XX 30-APR-1998; 98JP-0120550.
XX 30-APR-1998; 98JP-0120551.
XX 04-DEC-1998; 98JP-0345590.
XX 10-DEC-1998; 98JP-0351246.
XX
PA (SUMO ) SUMITOMO CHEM CO LTD.
XX
XX Watanabe E, Oeda K;
XX
XX WPI; 1999-593144/51.
XX DR N-PSDB; AAZ20210.
XX
XX New sense and antisense genes, useful for altering the level of
XX raffinose in food plants -
XX
PS Claim 27; Page 36-38; 55pp; English.
XX
XX This sequence represents rapeseed raffinose synthase, a protein
XX that can bind a D-galactosyl group through an alpha(1-6) bond to the
XX hydroxy group attached to the carbon atom at the 6-position of the
XX D-glucose residue in a sucrose molecule to form raffinose. cDNA
XX (see AAZ20210) encoding the enzyme was isolated from rapeseed cv.
XX Western leaf cDNA by PCR. Probes or primers generated from plant
XX raffinose synthase genes (see AAZ20207-10) may be used to obtain

```

```

CC other raffinose synthase genes by labeled detection or amplification
CC (claimed). These genes may be used to control the levels of
CC raffinose produced in plants. Antisense genes can be used to knock
CC out existing gene activity, and sense genes to increase the level
CC of gene activity. The resulting transgenic plants may be used as a
CC food source to alter the growing conditions for gut enterobacteria,
CC providing general health advantages.
XX
SQ Sequence 572 AA;
XX
Query Match 56.0%; Score 2364; DB 20; Length 572;
Best Local Similarity 73.9%; Pred. No. 6.7e-215;
Matches 430; Conservative 71; Mismatches 71; Indels 10; Gaps 4;
OY 203 LBEKTPRGIVDKFGWCTWAFILYVHPGVYIEGVRHLVDGCGCPGLVLLDDGQSGHNS 262
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 1 leektpgriivdkfgwctwafilyltpvdpghkvgkclvdgcpplvlddgwsglghds 60
OY 263 DPTTKEGMMQTVAGGEQMPRLKFOENYKFRDYVNPFRKATGPRAGOKMKAFIDELKGEFK 322
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 61 dgtldegmctvagegmprclpkrigenkfrdyvspk----dknevgmktatvrdleets 116
OY 323 TVEHVYVWHALGCGWGLRPQVGLPEARVIOVLSPLGLQMTMEDLAVDKIYLVKGLVP 382
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 117 tvdylyvwhalgcgywglrpgarlprrstlvrrpelspgjklmqdlavdkldtqgifs 176
OY 383 PEKAEEMVGLNHLKLEKVIDGVKIDVHLEMLCEDYCGRVDLAKAYVYKAMTKSTNKH 442
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 177 pdmanefegllshlnqvgingvkvdlhllemlckeygyrvdlakaykaltssvnhlf 236
OY 443 KGNQVITASMENCDPFLGTBAISIGRVGDDFMCTPSCDPPNGTFTMLQGCNHYHCAANDSL 502
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 237 dqnavaasmehcndlmfslgtaalslgyvddltwctapsdnglyvlgcmmhcaysn1 296
OY 503 WMGNFIRHPDWMFOSTHPCAFHAA5RAISGPRIVSDSVGKHNNDLKLVLPGSILR 562
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 297 wmgnfirhpdmfngsthpcaeethaa5raissgprivsdsvgkhnndlklvlpgs1lr 356
OY 563 SEYVALPTRDCLFEDPLNGETMLKIWNLNKFTGIVGAFNCGGGMCNRETRNOCESQYS 622
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 357 ceyvalptrdrilfeoplhngktmlk1wnlnkftgivaefnccgggcretrrdqctsgcv 416
OY 623 KRYTSKTPNKDIEHMGSENPISIEGVKTFALYVQAKKLILSKPSQDDIALDPFEFELI 682
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 417 nltatctpnndwsmgnpislenteeafalflsgskkilyvsqnddletlepfkfeil 476
OY 683 TVSPVTKLIQTSLHFAPIGLVNMNLNLSGALIOSVDYDDLSVTEIGVKGCGEMKVFASKRP 742
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 477 tvspvvtiegssvqtapiglvnmnlntsgalrslvlynee--sveilgyrgageltvyskkrp 534
OY 743 RACRIDGEDVGEFKYDDQWVVVQVPWIDSSGSGISVIEYLF 784
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 535 vackidgedvgefyg--eesmwmvqvpw---sapeglslskylf 572

RESULT 15
AAV70980
ID AAV70980 standard; Protein; 751 AA.
XX
XX
XX AAV70980;
XX
XX 09-AUG-2000 (first entry)
XX
XX Wheat raffinose synthase from clone wlm24.pk0021.n1.
XX
XX Wheat; raffinose synthase; raffinose saccharide; soybean;
XX clone wlm24.pk0021.n1; nutritional; soy protein.
XX
XX Triticum aestivum.
XX
XX MO200024915-A2.
XX

```

PD 04-MAY-2000.
XX
PF 22-OCT-1999; 99NO-US24923.
XX
PR 23-OCT-1998; 98US-0105451.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Allen SM, Hiltz WD;
XX
DR WPI; 2000-350754/30.
XX
DR N-PSDB; AAD00337.
XX
PT Nucleic acids and encoded proteins involved in the biosynthesis of
XX raffinose, useful for producing soybean seeds with a reduced raffinose
XX content and therefore improved nutritional quality -
PS
PS Claim 22; Page 52-54; 58pp; English.
XX
CC The present sequence is a raffinose synthase from
CC clone wlm24.pk0021.h1 isolated from a wheat seedlings cDNA
CC library wlm24. Raffinose synthase is involved in the biosynthesis
CC of raffinose and higher homologues in the raffinose saccharide family
CC from sucrose. The present sequence is useful for reducing the raffinose
CC saccharide content of soybean seeds which improves the nutritional
CC quality of the soy protein products derived from them.

SO Sequence 751 AA:

Query Match 35.3%; Score 1490; DB 21; Length 751;
Best Local Similarity 40.0%; Pred. No. 7.5e-132;
Matches 305; Conservative 132; Mismatches 270; Indels 56; Gaps 16;

QY 28 DGSFPTVNGHSLSDVDENTVYASPSPTSIDKSPVSGCFGFDASEPDSRHVYSIGIKL 87
DB 11 dgr-lavgrtvsvdpnvlaahasagajvd-----gatvgatageakshvftfglir 64
QY 88 DIRPMSTFRFKVWMTTHWVGNGDLESETQIVLE-----KSDGRRPYVFLPRIVE 139
DB 65 dcrfmclfrfklwmqtmqmgsgsrdrpletflllewpaaagnddsepylvmrlple 124
QY 140 GPFRTSTQPGDDDFVYCVESGSSKVVYDASFRSMYLHAGDPPALYKEMAKIVRTHLGT 199
DB 125 gffrtvlgndqdgllqcliesgdkavqlegmnsylnhaglnpfdltlgavkaveklmqt 184
QY 200 FRLEEKTRPRGIVKFGMCTDAFYLVHNGVTEGVNHLVGGCPRGLVLDGQMSIG 259
DB 185 fhhekkkvpstfvqwtgclwdaftydvtadgvykgjlrslaeggaaprfliddgwgqig 244
QY 260 HSDPITKEGNMOTV-AGEQMPCRLKFOENYKFRDYVNPRAKTPRAGQKGMKAIDELK 318
DB 245 sen-----kedpsvavgsaqafasrltqtkentkfqseqeet-----pqlkrlveetk 293
QY 319 GEFTVEHVYVHALCGYWGGLRPQVGLP--EARVIOPIVLSPLGIOMTMEDLAVDKIVLH 376
DB 294 ke-hgvksyvywamahagvygkvksaaemehesalavpyvsgpytgnqpdilvmdslsvl 352
QY 377 KVGIVPRPKAEEMTEGLHAHLEKVGIDGVKIDVYHLEMLCEDYGGGRVLDLAKAYKAMTK 436
DB 353 glglvhprkvystfydelhaylaacgvddvkvkdvqnlvelclgaingrayalcrayhralea 412
QY 437 SINHFNGNGVYASMEHNDPMFLGTEAISLGRYGDDEFWCTDPSGDPNGTFLWLGCHMYH 496
DB 413 svatnrfpngscismchntdmly-sakqlavvtasddtlyprdpas-----hcvh 460
QY 497 ---CANDSLMNGNFIHPDWMFOSTHPCAFHAASRAISSGPIYVSDSVGKHNFDLIKL 553
DB 461 issvayntlfigetmqpwmfnshlpaayhgaaraiygcpiyvskpghnfdallkkl 520
QY 554 VLPDGSILRSEYIALPTRDCLFEDPLHNGETMLKIMNLNKTGVIGAFNCOGGGRCRETR 613
DB 521 vlpdgsylraqljprtrcdlfsdpardgaallkikmmkcaayvgyvfncgagwcvvk 580

QY 614 RNOCFSOYSKRVTSKTNPKDIE--WHSGENPISIEGVKTFALYLIQAKKLIISKSQDL 670
DB 581 ktrihdeapgtlgsvraedvegltglddctgdav---vythragevlrlyprgatl 636
QY 671 DIALDPPEPEFLIYSPVTKLIQTSLHFAPIGLVNMLNTSGAIOV--DYDDDLSSVEIG 727
DB 637 pvtlkrleyelfhvcpy-ravapdisfapiglhlhmfnaagaveecvyrtneddkaavalr 695
QY 728 VKGCGEMRVFASKKPRACRIDGEDVGFERYDQD-MVVVQVWMP 769
DB 696 vrggrtgaycsrrpakcsidsadvefgydadtglyvtdvvpv 738

Search completed: November 30, 2001, 09:43:06
Job time: 165 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 30, 2001, 09:40:56 ; Search time 23.6 Seconds
(without alignments)
1218.018 Million cell updates/sec

Title: US-08-846-234-5
Perfect score: 4225
Sequence: 1 MARSPKNGSGNVVSPDGLND.....QVWPPIIDSSSGISVIEYLF 784

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 3664827 residues
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	115.5	2.7	902	1 YC47_SCHPO	014053 schizosacch
2	107	2.5	1174	1 CIEB_BACTA	003745 bacillus th
3	106	2.5	1723	1 A1M1_HUMAN	094741 homo sapien
4	105	2.5	730	1 GUGB_HAEIN	045177 haemophilus
5	102.5	2.4	564	1 HEMA_IARUD	019700 influenza a
6	101.5	2.4	478	1 FRBK_SALCH	000330 salmonella
7	101	2.4	514	1 V664_HSV11	000156 ictaluriid h
8	100.5	2.4	712	1 IRA1_HUMAN	051617 homo sapien
9	100	2.4	1550	1 GLTB_SYNY3	055037 synchocyst
10	99	2.3	770	1 GUGB_SYNY3	052981 synchocyst
11	99	2.3	1137	1 CYG1_CAEEL	008435 caenorhabdit
12	98	2.3	306	1 YCBU_BACSU	042242 bacillus su
13	97.5	2.3	1095	1 P1PA_DROME	013217 drosophila
14	97	2.3	520	1 HMCS_CRIGR	017304 cricetulus
15	97	2.3	520	1 HMCS_RAT	017425 rattus norv
16	97	2.3	859	1 ENV_EIAY2	022428 equine infe
17	97	2.3	1002	1 TAGA_VIRCH	024019 victrio chnol
18	97	2.3	3061	1 POLG_PVYHU	002963 p genome po
19	96.5	2.3	444	1 Y1H1_YEAST	040512 saccharomyc
20	96.5	2.3	631	1 DHM1_PARDE	012293 saccharomyc
21	96	2.3	460	1 ALN_YEAST	032375 saccharomyc
22	96	2.3	520	1 HMCS_HUMAN	001581 homo sapien
23	95.5	2.3	564	1 HEMA_TAGRE	019698 influenza a
24	95.5	2.3	964	1 KRPO_MCMV	011640 maize chlor
25	95	2.2	791	1 LON_CAMDE	069300 campylobact
26	94.5	2.2	733	1 AGAL_PEDPE	043467 pediococcus
27	94.5	2.2	971	1 NACL_RAT	001728 rattus norv
28	94	2.2	521	1 JMAL_XENLA	051370 xenopus lae
29	94	2.2	563	1 CHIA_SERMA	007254 serratia ma
30	94	2.2	887	1 CATE_RHINE	094576 rhizobium m
31	94	2.2	843	1 PULA_THEMA	038440 thebotoma
32	93	2.2	695	1 EFGI_SYNY3	028371 synchocyst
33	93	2.2	859	1 ENV_EIAY1	022427 equine infe

34	93	2.2	1047	1 ANPB_BOVIN	P46197 bos taurus
35	93	2.2	3988	1 POLG_BVDVN	P19711 bovine vira
36	92.5	2.2	400	1 EFTU_HERAV	P42477 herpesosiph
37	92.5	2.2	564	1 HEMA_IASE2	P19701 influenza a
38	92.5	2.2	1356	1 HET1_PODAN	000808 podospora a
39	92.5	2.2	1431	1 DAPK_HUMAN	053355 homo sapien
40	92	2.2	859	1 ENV_EIAYC	P32541 equine infe
41	92	2.2	966	1 M172_HUMAN	014596 homo sapien
42	92	2.2	1070	1 AGLU_CANTS	P28064 candida tsu
43	91.5	2.2	338	1 Y674_METUA	Q38087 methanococc
44	91.5	2.2	557	1 RSL_ECOLI	P02349 escherichia
45	91.5	2.2	1134	1 IF3X_HUMAN	O75153 homo sapien

ALIGNMENTS

RESULT 1	ID	YC47_SCHPO	STANDARD:	PRT:	902 AA.
AC	014053	YC47_SCHPO			
DT	15-JUL-1998	(Rel. 36, Created)			
DT	15-JUL-1998	(Rel. 36, Last sequence update)			
DT	20-AUG-2001	(Rel. 40, Last annotation update)			
DE	HYPOTHETICAL 100.6 KDA TRP-ASP REPEATS CONTAINING PROTEIN C1672.07 IN				
DE	CHROMOSOME III.				
GN	SPC1672.07.				
OS	Schizosaccharomyces pombe (fission yeast).				
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;				
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;				
OC	Schizosaccharomycetes.				
OX	NCBI_TaxID=4896;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=972;				
RA	Hilbert H., Duesterhoeft A., Wood V., Rajandream M.A., Barrell B.G.;				
RL	Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.				
CC	-I- SIMILARITY: CONTAINS 11 WD REPEATS (TRP-ASP DOMAINS).				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (see http://www.isb-sdb.ch/announce/				
CC	or send an email to license@sdb.ch).				
CC	-----				
DR	EMBL; AL031324; CAA20445.1; -				
DR	InterPro: IPR001680; WD40.				
DR	Pfam: PF00400; WD40; 8				
DR	SMART; SM00320; WD40; 5				
DR	PROSITE; PS00678; WD_REPEATS_1; 3				
DR	PROSITE; PS50082; WD_REPEATS_2; 2				
DR	PROSITE; PS50294; WD_REPEATS_REGION; 1				
KW	Hypothetical protein; Repeat; WD repeat.				
FT	REPEAT 31	71			
FT	REPEAT 111	150			
FT	REPEAT 158	198			
FT	REPEAT 202	241			
FT	REPEAT 243	283			
FT	REPEAT 287	328			
FT	REPEAT 333	376			
FT	REPEAT 393	437			
FT	REPEAT 487	528			
FT	REPEAT 570	611			
FT	REPEAT 613	651			
SO	SEQUENCE	902 AA;	100569 MW;	SEEFEC6034BDC047 CRC64;	

Query Match 2.7%; Score 115.5; DB 1; Length 902;
Best Local Similarity 20.0%; Pred. No. 0.31;
Matches 139; Conservative 85; Mismatches 231; Indels 241; Gaps 36;


```

DB 242 RISFWYIVEPGDLIVFNTIGNLIAPRGHYKLNSOKKSTILNTAVPISGCVSKCHTNRGS 301
OY 164 KVVVDSFBRMLVLHGGDDDFALVKKEMKIVRTHLGFRLLLEETPPGICYDKKGMCTWDAF 223
DB 302 ITTTPFPONISRTISGDCRKYVKGSLKLTAT---GMRNIPKANTRGTP----- 346
OY 224 YLTVHPQVIEGVRHNLVDGCPGVLVIDDGNOSIGHSDSPITKESGMOTVAGEQMPCHRL 283
DB 347 -----GAIAG-----FIENGMOGL----- 360
OY 284 LKFOENYKFRDYVNPKATPRPAGQKMAFIDELKGEFTEVHYVMAHLCGYGLRPQ 343
DB 361 --IDGWYGER-IQNEAGTSTAADLKSTQAIIDQINGKLNRL----- 398
OY 344 VGLPFRARYIOPVLSGLQMTMEDLAVDKIVLHKVGLVPEKAEEMEGHIAHLEKVGID 403
DB 399 -----IEKTNEK-----YHQI-----EKEFEQVEERIODLEKY-VE 428
OY 404 GVKRIDV-----IHLEMLCEDYGGVLDLAKAYYKAMTKSI-----NKHFKGNGVIASMEH 453
DB 429 DTKIDLMSTNAELVALLENQH--TIDVTSEMKKLFERYRQRLRENAEDKNGCFELFHD 486
OY 454 CNDPFLGTSAISLGRVGDFFWCTDPSGDPNGTFWLOGCHMYHCAND-SLWM 504
DB 487 CDNNCI---ESINGTYDHDYHDEAI---NNRFQIGVYKLTGQYKDIILMI 532

RESULT 6
REFB_SALCH STANDARD; PRT; 478 AA.
ID 000330:
AC 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE PHOSPHOMANNOMUTASE (EC 5.4.2.8) (PMM).
GN MANB OR RFBK.
OS Salmonella cholerae-suis (Salmonella enterica).
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=591;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M67;
RX MEDLINE=92349966; PubMed=1379320;
RA Brown P.K., Romana L.K., Reeves P.R.;
RT Molecular analysis of the rfb gene cluster of Salmonella serovar
RT muenchen (strain M67): the genetic basis of the polymorphism between
RT groups C2 and B."
RT Mol. Microbiol. 6:1385-1394(1992).
RL - FUNCTION: INVOLVED IN GDP-MANNOSE BIOSYNTHESIS WHICH SERVES AS
CC THE ACTIVATED SUGAR NUCLEOTIDE PRECURSOR FOR MANNOSE RESIDUES
CC IN CELL SURFACE POLYSACCHARIDES. THIS ENZYME PARTICIPATES IN
CC SYNTHESIS OF THE LPS GROUP C2 O ANTIGEN.
CC -1- CATALYTIC ACTIVITY: D-MANNOSE 1-PHOSPHATE -> D-MANNOSE 6-PHOSPHATE.
CC -1- PATHWAY: GDP-MANNOSE BIOSYNTHESIS WITHIN THE O ANTIGEN
CC BIOSYNTHESIS PATHWAY OF LIPOPOLYSACCHARIDE BIOSYNTHESIS.
CC -1- SIMILARITY: BELONGS TO THE PHOSPHOHEXOSE MUTASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL: X61917; CAA43916.1; -
DB PIR: S22622; S22622.
DB InterPro: IPR001485; PGM_PMM.
DB Pfam: PF00408; PGM_PMM; 1.
DB PROSITE: PS00710; PGM_PMM; 1.
KW LIPOPOLYSACCHARIDE BIOSYNTHESIS; Isomerase; Transmembrane;
KW Phosphorylation.

```

```

FT TRANSMEM 30 46 POTENTIAL.
FT TRANSMEM 265 284 POTENTIAL.
FT ACT_SITE 111 111 FORMS THE PHOSPHOERINE INTERMEDIATE
FT (BY SIMILARITY).
SQ SEQUENCE 478 AA; 52813 MW; 982E5B85083B893 CRC64;

Query Match
Best Local Similarity 2.4%; Score 101.5; DB 1; Length 478;
Matches 59; Conservative 30; Mismatches 79; Indels 103; Gaps 14;

OY 349 EARVIOPLVSPGLQMTMEDLAVDKIVLHKVGLVPEKAEEMEGHIAHLEKVGIDKID 408
DB 141 DASFMQPKLE---QLTISTIAARNYILRYTSLFPMPEFLNKKRIGIYEH----- 185
OY 409 VIHLEMLCEDYGGVLDLAKAYYKAMTKSINKHFKNGVIASHEHNDPFLCTEASISG 468
DB 186 -----SSAGR-DLYKTFKML-----GATVYLSARDERVIDTEAVS-- 222
OY 469 RVGDD-----FMCTDPSG-----DPNGTFWLOG-----CHM----- 494
DB 223 --EDDKNKAITNAKKYQOLAIFSTDGDRPLADRYGN-WLRGDLILGLSLELAADAV 279
OY 495 ---VHCANDSLWNGNFI-HPDWMFQSTHPCAFHAAASRAISGGPIYSDSVGKNHFDLL 550
DB 280 AIPVSC-NSTISGNGFKEFKHYERTKISPYIAFAKLSA-----NYNCI 322
OY 551 KKLVPDGSILNSEY-----ALPTROCL 574
DB 323 AGFEANGFLLGSDVYINORLLAKLPTRDAL 353

RESULT 7
ID V664_HSV11 STANDARD; PRT; 514 AA.
AC 000156:
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DE 01-DEC-1992 (Rel. 24, Last annotation update)
DE HYPOTHETICAL GENE 64 PROTEIN.
GN 64.
OS Ictalurid herpesvirus 1 (Channel catfish virus) (CCV).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC unclassified Herpesviridae.
OX NCBI_TaxID=10401;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AUBURN 1;
RX MEDLINE=92087490; PubMed=1727613;
RA Davidson A.J.;
RT "Channel catfish virus: a new type of herpesvirus.";
RT Virology 186:9-14(1992).
RL -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL: M75136; AAA88168.1; -
DB PIR: A36793; A36793.
KW Hypothetical protein.
SQ SEQUENCE 514 AA; 58198 MW; 88E7302A1424D29C CRC64;

Query Match
Best Local Similarity 2.4%; Score 101; DB 1; Length 514;
Matches 108; Conservative 75; Mismatches 177; Indels 172; Gaps 28;

OY 33 TVNGHSFLSDVPENIVASPSPTYSIDKSPVSGCFVG-----FDASEPDSRHVYSI 83
DB 111

```

```

Db 44 TLEGEDEGCIQELINLVLYEKMAKSTSELSGCVTYALMTLGIFFDR-----SL 96
Oy 84 GKLKDIRFMSIFRFKVMWTTTHWYGRNGDLESETOYILEKSDGRPYVFL---PIVGG 140
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 97 KRLVDSGVLGMRFRNMLRVFFGVATGD-----DALIGLIDTWFAPVILARFVIDA 151
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 141 PRTSIQPDGDDFVDCV-ESGSS-KVDAFSRSMYLH-----AGDDP-----FAL 185
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 152 -----QLIDSNFCVLCMYESHSLATADTRFETIICDHFTASVYAGGPPCSAEVAV 205
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 186 VKEMAKIVTHLTGTFRLLEKTPRGIVDFGMCWTMAFYLTVAPGVISGRHLVDGGCP 245
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 206 VKKTYQI-DLSLDF---HERVARG-----DVAKEL-----232
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 246 PGLVLIDDMOSIGHSDSDITKEG-----MNOVAGEOMPCLRLKPEQNYKFR 293
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 233 -----AVKRPJTTLTRKNGSINNPTKALNRAREFCANDRESSLVEYQKRLREV 281
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 294 DYVNPRAATGPRAGQCKMAFIDELKGEFTEVEHYVWHALCGWGLRPOVPLPEARVI 353
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 282 NNNINLTNP---KNLNTFAELFNVFVHAH-----RRKTI 315
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 354 OPLSPGL-----QMTMEDLAVDKIVLHKV-----GLVPEKAEEMYEGLHA 395
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 316 RERLSDGLGDMKGRRAHMLFTTKRLTAVDKNLLTITTFYAAAPRYRYNKKYNNHA 375
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 396 -----HLEKVGI-DGV-----KIDVILHLEMLCEDYGGRVDLAKAYYKAMTKSINKH 441
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 376 HRVIFPKHMSIGFTNEGIFAFQYOLNOVDLREVGCODY--RSIDIPIITVNA--ENFNRL 431
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 442 FKNGCYIASHEHNDFMFLCTEASISLGRVDDDWPCDPSD--PNTGFWLOG 491
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 432 TELEWYIANI--CT-FIFHNKTKIKLH-GD---APDLNIDELPDGLYLFNG 476
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 8
IRAL_HUMAN STANDARD; PRT; 712 AA.
AC P51617;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE INTERLEUKIN-1 RECEPTOR-ASSOCIATED KINASE 1 (EC 2.7.1.1-) (IRAK-1).
GN IRAK1 OR IRAK.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=96180673; PubMed=8599092;
RA Cao Z., Henzel W.J., Gao X.;
RT "IRAK: a kinase associated with the interleukin-1 receptor.";
RL Science 271:1128-1131(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Platzer M., Bauer D., Drescher B.;
RN Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RA Reichwald K., Kioschis P., Rosenthal A., Platzer M.;
RN Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INVOLVED IN IL-1 PATHWAY. THIS KINASE ASSOCIATES WITH
CC THE IL-1 RECEPTOR IL-1R-1. THIS ASSOCIATION IS RAPID AND IL-1
CC DEPENDENT.
CC -1- TISSUE SPECIFICITY: SEEMS TO BE UBIQUITOUS, ALTHOUGH PRESENT IN
CC SMALL AMOUNTS.
CC -1- PTM: AN EXTENSIVE PHOSPHORYLATION OF IRAK OCCURS AFTER ITS
CC ASSOCIATION WITH IL-1R-1. THIS STEP COULD BE LINKED TO THE
CC ACTIVATION OF THE KINASE.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC BELLE SUBFAMILY.
CC -----

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-stb.ch/announce/
CC or send an email to license@isb-stb.ch).
CC -----
CC EMBL: L76191; AAC1949.1; -
CC EMBL: U52112; AAC51752.1; -
CC EMBL: AF030876; AAC08756.1; -
CC EMBL: AF031075; AAF21636.1; -
CC MIM: 300283; -
CC InterPro: IPR000719; Euk_pkinase.
CC InterPro: IPR002965; P_rich_extensn.
CC InterPro: IPR002290; Ser_thr_kin_actsite.
CC Pfam: PF00069; pkinase; 2.
CC PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
CC PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
CC Transferrase; Serine/threonine-protein kinase; ATP-binding.
CC DOMAIN 212 521
CC NP_BIND 218 226 ATP (BY SIMILARITY).
CC FT BINDING 239 239 ATP (BY SIMILARITY).
CC FT ACT_SITE 338 338 BY SIMILARITY.
CC FT CONFLICT 196 196 F -> S (IN REF. 1).
CC FT CONFLICT 532 532 S -> L (IN REF. 1).
CC FT SEQUENCE 712 AA; 76536 MW; A7ADED75D3A3961D CRC64;

Query Match 2.4%; Score 100.5; DB 1; Length 712;
Best Local Similarity 27.0%; Pred. No. 3.7;
Matches 38; Conservative 17; Mismatches 61; Indels 25; Gaps 6;

Oy 116 ETQIVILEKSDSGRPYVFLPIVEGPRISIQPDDEFVDCV-ESGSKVYDASFRSM 173
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 184 ESSVSLQ---GARPPFCPLCE-----ISRGTHNFSEELKIGEGFGCVYRAVWRNT 234
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 174 LYL-----HAGDPEFALYKAMKIVRTHLCTFRLLEKTPRGIVDFGMCWTMAFYLTVH 228
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 235 VYAVKRLKENADLEWVAVKQSFLEVEQLSRF-----HPNIVDFAGYCAONGFFCLVY 288
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 229 ---PGVIEGVRLVDGCGPP 246
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 289 GELPNSGLEDRHLCQYQACPP 309
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 9
GLTB_SYNY3 STANDARD; PRT; 1550 AA.
ID GLTB_SYNY3
AC P55037;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE FERREDOXIN-DEPENDENT GLUTAMATE SYNTHASE 1 (EC 1.4.7.1) (FD-GOGAT).
GN GLTB OR SLI1502.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95244836; PubMed=7727752;
RA Navarro F., Chavez S., Candau P., Florencio F.J.;
RT "Existence of two ferredoxin-glutamate synthases in the
RT cyanobacterium Synechocystis sp. PCC 6803. Isolation and insertion
RT inactivation of gltb and glts genes.";
RL Plant Mol. Biol. 27:753-767(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirosewa M., Sugiyura M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Nartuo K.,

```

RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
 RA Yamada M., Yasuda M., Tabata S.;
 RT *Sequence analysis of the genome of the unicellular cyanobacterium
 RT *Synechocystis* sp. strain PCC6803. II. Sequence determination of the
 RT entire genome and assignment of potential protein-coding regions.";
 RL DNA Res. 3:109-136(1996).
 CC -1- CATALYTIC ACTIVITY: 2 L-GLUTAMATE + 2 OXIDIZED FERREDOXIN -
 CC L-GLUTAMINE + 2 OXOGLOUTARATE + 2 REDUCED FERREDOXIN.
 CC -1- COFACTOR: BINDS A 3FE-4S CLUSTER. FAD AND FMN.
 CC -1- PATHWAY: GLUTAMINE SYNTHETASE/COGAT PATHWAY WHICH IS INVOLVED
 CC IN THE ASSIMILATION OF AMMONIA.
 CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-21 IS THE INITIATOR.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: X80485; CAAS6652.1; -
 CC EMBL: D90902; BAA17018.1; -
 CC HSSP: P02392; ICTF.
 CC InterPro: IPR002489; DUF14.
 CC InterPro: IPR002932; GLU_synthase.
 CC Pfam: PF01493; DUF14; 1.
 CC Pfam: PF01645; Glu_synthase; 1.
 CC Oxidoreductase; Iron-sulfur; 3Fe-4S; Flavoprotein; FAD; FMN;
 CC Glutamate biosynthesis; Complete proteome.
 CC FMN (BY SIMILARITY).
 CC FT DOMAIN 43 393
 CC FT METAL 1097 1154
 CC FT METAL 1150 1150 IRON-SULFUR (3FE-4S) (BY SIMILARITY).
 CC FT METAL 1156 1156 IRON-SULFUR (3FE-4S) (BY SIMILARITY).
 CC FT METAL 1161 1161 IRON-SULFUR (3FE-4S) (BY SIMILARITY).
 CC SEQUENCE 1550 AA; 169071 MW; 07AA64D832296943 CRC64;
 Query Match 2.48; Score 100; DB 1; Length 1550;
 Best Local Similarity 19.9%; Pred. No. 13;
 Matches 136; Conservative 89; Mismatches 224; Indels 234; Gaps 38;
 QY 120 VLEKSDSGRPVFLPVGSPRTS-----IOPGDDDFVGVCESSSKVVDASFRSMLY 175
 DB 406 LVTLASAG-----VLPY--EPERVAKKGRQDGRMFLVD--MEQGRV--IADEIK----- 451
 QY 176 LHAQDDPFLVLEKMKIVRTN-----LGFRLDEKTPPGIVD-----RFGW 217
 DB 452 -----QELVSOHPYCEMLANLKSLEQLPSPGVNPTDAESLRQRMAFGY 497
 QY 218 CTMDAFYLVNPOGVIEGVNHLVDGCGPGLVLLIDG-----WOSIGHSDPI 265
 DB 498 -TFEELRLILAPMG-RDGVSAIGSGADPLAVLSDKPKLLNYFOQLFAQVTPRDISI 555
 QY 266 TRE---GMNOTVAGE-----QMPRL-----LKEQENKFRRYVPA 300
 DB 556 REELITSAETTIGEGNMLDPRPESCRLELKPRLTNEDLAKALADDEFKSVTLIDL 615
 QY 301 TGRPAGOGKMAFIDELKGEFK-----TVEHYVNHALCGYWG----- 338
 DB 616 FDPNOCGAGLKTALDNLFTENDQAIISOGANLILSDROVSEKAAIT-PALLAVSGLLNHL 674
 QY 339 ---GLRPOVGLPEARVIOPLVSPGLQMTMEDLAVDKITVLKVLGVLPPEKAEMYEGILHA 395
 DB 675 IRNRSRTKV-GL-----VLESGBREVHNFV--LLGYGCAINPYLAFTLIDGMA 723
 QY 396 HLEAVGIDGVKIVDILHLEMLCEYDGGVRLDAKAYYKAMTKSTIKHKGNCVISMENCN 455
 DB 724 -----EGLVNVNDH--KTAC-----KNYIKAAKRGVIAKVASIGISTIQSTRG 764
 QY 456 DFMELTEATISLGR-VGDDDFWCCTDPSGDNGCTFVYLGCHVHNCANDSLAMGNFIHPDMD 514

DB 765 AQIF---EAVGLNQSVIDEYFCHRTSSR-----IOGSDLGVIQAEALR----- 804
 QY 515 FOSTHPCAFHNASRAISGSPVIVSDYVGKHN-----DLKKLVLPDGS 560
 DB 805 -----HQAFAPRPDLTLTDGGERQVRKQDEEHLFSPOTIHLQRAVRGNYE 854
 QY 561 LRSEYVAL-----PTPDCV-FED-----PLHNETMLKINLNF-TGVIGAFNCG 605
 DB 855 LVQYALVALVEQNGKFFTLGLDPLDRESITPLEVEPEIAL--MKRFKGAMSY----- 907
 QY 606 GGCREFTRNNQCSQSKRYTSKTN-----PKDIEMHSGENPISIEGKTA----- 652
 DB 908 GSISKKEHESLAIAM--NRGKGSNTGEGEDPERFTWTDGDSKNAIKOVASGRFGV 965
 QY 653 --LYIQAOKLIL-----SKPSQ 668
 DB 966 TSLYLSQAKETIQIMAGAKPGE 988
 RESULT 10
 GLGB_SYNY3
 ID GLGB_SYNY3 STANDARD; PRT; 770 AA.
 AC P52961;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE 1.4-ALPHA-GLUCAN BRANCHING ENZYME (EC 2.4.1.18) (GLYCOGEN BRANCHING
 DE ENZYME).
 GN GLGB OR SLU0158.
 OS *Synechocystis* sp. (strain PCC 6803).
 OX Bacteria; Cyanobacteria; Chroococcales; *Synechocystis*.
 OX NCBI_Taxid=1148;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96127529; PubMed=8590279;
 RA Kaneo T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
 RA Sugita M., Tabata S.;
 RT *Sequence analysis of the genome of the unicellular cyanobacterium
 RT *Synechocystis* sp. strain PCC6803. I. Sequence features in the 1 Mb
 RT region from map positions 648 to 928 of the genome.";
 RL DNA Res. 2:153-166(1995).
 CC -1- CATALYTIC ACTIVITY: FORMATION OF 1,6-GLUCOSIDIC LINKAGES OF
 CC GLYCOGEN.
 CC -1- PATHWAY: THIRD STEP IN GLYCOGEN BIOSYNTHESIS.
 CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
 CC KNOWN AS THE ALPHA-AMYLASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: D63999; BAA10073.1; -
 CC InterPro: IPR000461; Alpha_amylase.
 CC Pfam: PF00128; alpha-amylase; 1.
 CC KM Glycogen biosynthesis; Transferase; Glycosyltransferase;
 CC Complete proteome.
 CC FT ACT_SITE 433 433 BY SIMILARITY.
 CC FT ACT_SITE 486 486 BY SIMILARITY.
 CC FT ACT_SITE 554 554 BY SIMILARITY.
 CC SEQUENCE 770 AA; 89527 MW; A435AFCA7703FA8A CRC64;
 Query Match 2.38; Score 99; DB 1; Length 770;
 Best Local Similarity 18.08; Pred. No. 5.4; Indels 160; Gaps 24;
 Matches 77; Conservative 63; Mismatches 128; Indels 160; Gaps 24;
 QY 205 EKPSPGVLDKFGMCTWDAFYLTVHPQ--GVIEGVNHLVDGCGPGLVLLIDGMSIGHS 262

Db 1026 RM 1027

RESULT 12

ID YCBJ_BACSU STANDARD: PRT: 306 AA.

AC P42242:

DT 01-NOV-1995 (Rel. 32, Created)

DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE HYPOTHETICAL 34.5 KDA PROTEIN IN GLTF-CWLJ INTERGENIC REGION (ORF9).

GN YCBJ.

OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

CC Bacillus/Staphylococcus group; Bacillus.

OK NCBI_TaxID=1423;

RP SEQUENCE OF 98-306 FROM N.A.

RC STRAIN-168:

RX MEDLINE=95219079; PubMed=7704254;

RA Ogawa K.-I., Akagawa E., Nakamura K., Yamane K.;

RT "determination of a 21348 bp nucleotide sequence around the 24 degrees region of the Bacillus subtilis chromosome.";

RL Microbiology 141:269-275(1995).

RM [2]

RP SEQUENCE FROM N.A.

RC STRAIN-168:

RA Kunst F., Ogawa K.-I., Yoshikawa H., Danchin A.;

RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: TO AMINOGLYCOSIDE PHOSPHOTRANSFERASES AND SPECIFICALLY TO VIOMYCIN PHOSPHOTRANSFERASE.

CC -1- SIMILARITY: TO AMINOGLYCOSIDE PHOSPHOTRANSFERASES AND SPECIFICALLY TO VIOMYCIN PHOSPHOTRANSFERASE.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: D30808; BAA06474.1; .

DR EMBL: 299105; CAB12046.1; .

DR Subtilisin; Bg11165; ycbj.

KM Hypothetical protein; Transferase; Kinase; ATP-binding;

KM Complete proteome.

FT ACT SITE 204 BY SIMILARITY.

FT SEQUENCE 306 AA; 34481 MW; 5EF0FD9EC085977D CRC64;

SO

Query Match 2.3%; Score 98; DB 1; Length 306;

Best Local Similarity 20.1%; Pred. No. 1.7;

Matches 54; Conservative 33; Mismatches 84; Indels 98; Gaps 11;

Db 120 VILEKSDSGRPV-----FL-----LPIV 138

Db 40 VVEKADDTGMPVYLKKRSDYVERASAGITLAFRLANLTDVDPWRHTPELLAYPML 99

Db 139 EEPFRTST-----QPGDDPY-----DYCVESGSSKVVYDASRSMKLYLHAGD 181

Db 100 KCTPAAGIDLEKQYVWMMNDHPSPDDEFRLADILAE-----LH-GTD 142

Db 182 PALTVEAMKIVRTGLPRLLEKTPPGIVDKFGWCT--WDPAFLTVHPGVIEGVRIL 239

Db 143 QISAGSGGLEVRPE--DFRQMTADSMVDVKNKLGSTLMEKMWKVVDDAYWPGFSSL 200

Db 240 YDGGCPGLVLLIDDSMOSIG-----HSDPITTEKGMQTVAGGDMPCRLKFOENYKF 292

Db 201 IKGDLHPHILLIDQNGRVYGLDWTETAKVADPAKDFVLYQTIFGEKTRALLEIYDQ----- 257

Db 293 KDYVNPKATGPRAGOKMKAFIDELKGEF 321

Db 258 -----AGGRIMWK-MOEHISEMOAAV 277

RESULT 13

ID PIPA_DROME STANDARD: PRT: 1095 AA.

AC P13217;

DT 01-JAN-1990 (Rel. 13, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE 1-PROSPHATIDYLINOSITOL-4,5-BISPHOSPHATE PHOSPHODIESTERASE

DE (EC 3.1.4.11) (PHOSPHOTRANSFERASE-SPECIFIC PHOSPHOLIPASE C).

GN NORPA.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

CC Ephydroidea; Drosophilidae; Drosophila.

OK NCBI_TaxID=7227;

RP SEQUENCE FROM N.A.

RC MEDLINE=8831074; PubMed=2457447;

RA Bloomquist B.T., Shortridge R.D., Schneuwly S., Perdev M.H.,

RT Montell C., Steller H., Rudin G., Pak W.L.;

RT "Isolation of a putative phospholipase C gene of Drosophila, norpa, and its role in phototransduction.";

RL Cell 54:723-733(1988).

CC -1- FUNCTION: THE PRODUCTION OF THE SECOND MESSENGER MOLECULES DIACYLGLYCEROL (DAG) AND INOSITOL 1,4,5-TRISPHOSPHATE (IP3) IS MEDIATED BY ACTIVATED PHOSPHATIDYLINOSITOL-SPECIFIC PHOSPHOLIPASE C ENZYMES. NORPA IS INVOLVED IN PHOTOTRANSDUCTION.

CC -1- CATALYTIC ACTIVITY: 1-PHOSPHATIDYL-D-MYO-INOSITOL 4,5-BISPHOSPHATE + H(2)O = D-MYO-INOSITOL 1,4,5-TRISPHOSPHATE + DIACYLGLYCEROL.

CC -1- SIMILARITY: DOMAINS X AND Y ARE CONSERVED IN DIFFERENT FORMS OF PLC AND ARE ESSENTIAL FOR CATALYTIC ACTIVITY.

CC -1- SIMILARITY: CONTAINS 1 C2 DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: J03138; AAA28724.1; .

DR PIR: A31225; A31225.

DR HSSP: P10688; 10AS.

DR FlyBase; FBgn004625; norpa.

DR InterPro; IPR000008; C2.

DR InterPro; IPR001192; PL-PLC.

DR InterPro; IPR000909; PL-PLC_X.

DR InterPro; IPR001711; PL-PLC_Y.

DR Pfam; PF00168; C2; 1.

DR Pfam; PF00388; PI-PLC-X; 1.

DR Pfam; PF00388; PI-PLC-Y; 1.

DR PRINTS; PR00390; PPHPLIPASEC.

DR PRODOM; PD001202; PL-PLC_Y; 1.

DR SMART; SM00239; C2; 1.

DR SMART; SM00148; PLCX; 1.

DR SMART; SM00149; PLCY; 1.

DR PROSITE; PS50004; C2_DOMAIN_2; 1.

DR PROSITE; PS50007; PIPLC_X_DOMAIN; 1.

DR PROSITE; PS50008; PIPLC_Y_DOMAIN; 1.

KM Hydrolyase; Lipid degradation; Vision; Transducer.

FT DOMAIN 320 469 DOMAIN X.

FT DOMAIN 530 666 DOMAIN Y.

FT DOMAIN 673 771 C2 DOMAIN.

FT ACT SITE 334 334 BY SIMILARITY.

FT ACT SITE 381 381 BY SIMILARITY.

FT SEQUENCE 1095 AA; 124822 MW; 2D945EF0ACBE69B3 CRC64;

SO

Query Match 2.3%; Score 97.5; DB 1; Length 1095;

Best Local Similarity 19.4%; Pred. No. 12;

DE SYNTHASE) (3-HYDROXY-3-METHYLGUTARYL COENZYME A SYNTHASE).
GN HMGCS1 OR HMGCS.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Liver;
RX MEDLINE=90301491; PubMed=1972979;
RA Ayle J., Gil-Gomez G., Hegardt F.G.;
RT "Nucleotide sequence of a rat liver cDNA encoding the cytosolic 3-
hydroxy-3-methylglutaryl coenzyme A synthase.";
RL Nucleic Acids Res. 18:3642-3642(1990).
CC -1- FUNCTION: THIS ENZYME CONDENSES ACETYL-COA WITH ACETOACETYL-COA
TO FORM HMG-COA, WHICH IS THE SUBSTRATE FOR HMG-COA REDUCTASE.
CC -1- CATALYTIC ACTIVITY: (S)-3-HYDROXY-3-METHYLGUTARYL-COA + COA -
ACETYL-COA + H(2)O + ACETOACETYL-COA.
CC -1- PATHWAY: PRODUCTION OF MEVALONATE FROM HMG-COA PRIOR TO THE
SYNTHESIS OF STEROLS SUCH AS CHOLESTEROL AND ISOPRENOIDS.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE HMG-COA SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X52625; CAA36852.1; .
DR PIR: S12736; S12736.
DR InterPro: IPR000590; HMG_COA_synth.
DR Pfam: PF01154; HMG_COA_synth. 1.
DR PROSITE: PS01226; HMG_COA_SYNTHASE; 1.
KW Lyase; Cholesterol biosynthesis; Multigene family.
FT ACT_SITE 129 129 POTENTIAL.
SQ SEQUENCE 520 AA; 57433 MW; CB213A27B0C177CB CRC64;

Query Match 2.3%; Score 97; DB 1; Length 520;

Best Local Similarity 18.8%; Pred. No. 4.4; Mismatches 117; Indels 166; Gaps 21;

Matches 80; Conservative 62; Mismatches 117; Indels 166; Gaps 21;

OY 378 VGLVPE-----KAE-EMEGLHALEKVGIDGKIDYIHLEMCEDYGRVDA 427
DB 17 VGIVALEIYFQYVQDALEKRYDGVDAKRYTIGQAAMG-----FCTD--REDIN 66
OY 428 KAYKAMTRKSIKHKNGVISMHCNDFMFTGEAI-----SLGRVGDDEWCTD 478
DB 67 SLCLTVVOKLMERNLSLSDICIRLE-----VGTETITDKSKSVKSNLMQLPEESGNTD 119
OY 479 PEGDP-----NGTFMLOGCHVHCANDSLWGNF--IHPDMFQS--THP 520
DB 120 IEGIDTNAACYGTAFAVFAVMIE-----SSMDGRALVYAGDIATYASGNAP 170
OY 521 CAFFHAASRAISGGPIYSD-----SVCKHNFDLKLVLPDGSILRSEYVALPTRDCLP 575
DB 171 TGGVAGVALLIPNAPVIFDRGLRGTIMOHAYFYK---PD---MLSEY----- 213
OY 576 EDDLNGETMLKIWNLKNKTGIVGAFNCGGGMCHRETRNOGCFOSYKRVTSKTPKDI 635
DB 214 --PVYDGKLSIO-----CYLSALDRCTSVYRKIRI-----Q 243
OY 636 WHSGENPISIEGVKTFALYLOAKKLLIKSPQDLIDALPPEFELITVSPYTKLIOTSL 695
DB 244 W-----QKEGKDKDFTLNDFGF--MIFHSPYCKLVOKSL 275
OY 696 HPAPIGLVNMLTSGAIGSVYDDLSVIEIGKGGEMRVFASKPRACRIDGEDVGRK 755
DB 276 --ARWFLNDFLN-----DONRDKNSIYSGLEAFGDVKL-----EDTYFD 312

OY 756 YDODQ 760
DB 313 RDYER 317

Search completed: November 30, 2001, 09:41:02
Job time: 41:sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 30, 2001, 09:43:39 ; Search time 34.66 seconds
(without alignments)
509.018 Million cell updates/sec

Title: US-08-846-234-5
Perfect score: 4225
Sequence: 1 MAPSFKNNGSNVVSFDGLND.....QVPWPIDSSSGISVIEYLF 784

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Dackfilest1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4225	100.0	784	US-08-846-234-5	Sequence 5, Appl1
2	181	4.3	30	US-08-846-234-1	Sequence 1, Appl1
3	109	2.6	1390	US-08-770-544-2	Sequence 2, Appl1
4	107.5	2.5	1584	US-09-251-645-6	Sequence 6, Appl1
5	107	2.5	1174	US-08-040-751-3	Sequence 3, Appl1
6	107	2.5	1174	US-08-291-368-2	Sequence 2, Appl1
7	107	2.5	1174	US-08-962-190-2	Sequence 2, Appl1
8	107	2.5	1174	PCR-US95-10310-2	Sequence 2, Appl1
9	107	2.5	1174	5164180-4	Patent No. 5164180
10	105	2.5	19	US-08-846-234-2	Sequence 2, Appl1
11	100	2.4	520	US-08-305-505-6	Sequence 6, Appl1
12	100	2.4	1276	US-08-717-515-8	Sequence 8, Appl1
13	98.5	2.3	770	US-08-525-654A-1	Sequence 1, Appl1
14	98.5	2.3	771	US-08-525-654A-3	Sequence 3, Appl1
15	97	2.3	520	US-08-305-505-4	Sequence 4, Appl1
16	97	2.3	520	US-08-305-505-5	Sequence 5, Appl1
17	95.5	2.3	712	US-08-587-889-2	Sequence 2, Appl1
18	95.5	2.3	712	US-08-980-060-5	Sequence 5, Appl1
19	95.5	2.3	712	US-09-307-185-5	Sequence 5, Appl1
20	95.5	2.3	712	PCR-US96-09193-2	Sequence 2, Appl1
21	94	2.2	750	US-09-364-970-8	Sequence 8, Appl1
22	93.5	2.2	540	US-08-687-580B-7	Sequence 7, Appl1
23	92.5	2.2	1423	US-08-810-712-10	Sequence 10, Appl1
24	92	2.2	280	US-08-403-853-10	Sequence 10, Appl1
25	92	2.2	649	US-09-066-047-5	Sequence 5, Appl1
26	92	2.2	820	US-09-313-677-21	Sequence 21, Appl1
27	92	2.2	926	US-09-313-677-2	Sequence 2, Appl1

28	92	2.2	933	4	US-09-313-677-19	Sequence 19, Appl1
29	92	2.2	967	4	US-09-313-677-17	Sequence 17, Appl1
30	89	2.1	559	4	US-09-242-690A-15	Sequence 15, Appl1
31	89	2.1	609	3	US-08-941-445A-7	Sequence 7, Appl1
32	88.5	2.1	2547	3	US-09-058-489-35	Sequence 35, Appl1
33	88	2.1	842	2	US-08-928-692-18	Sequence 18, Appl1
34	88	2.1	3165	2	US-08-459-146-3	Sequence 3, Appl1
35	88	2.1	3165	2	US-08-459-065-3	Sequence 3, Appl1
36	87.5	2.1	876	1	US-08-717-515-4	Sequence 4, Appl1
37	87.5	2.1	908	2	US-08-249-380-2	Sequence 2, Appl1
38	87.5	2.1	1410	3	US-09-335-409-3	Sequence 3, Appl1
39	87	2.1	287	2	US-08-424-641B-10	Sequence 10, Appl1
40	87	2.1	287	2	US-08-820-980-10	Sequence 10, Appl1
41	87	2.1	287	2	US-08-826-439-10	Sequence 10, Appl1
42	87	2.1	325	4	US-08-961-083-80	Sequence 80, Appl1
43	87	2.1	11877	4	US-09-105-537-6	Sequence 6, Appl1
44	86.5	2.0	627	4	US-09-273-163-4	Sequence 4, Appl1
45	86	2.0	677	4	US-08-836-567-2	Sequence 2, Appl1

ALIGNMENTS

```
RESULT 1
US-08-846-234-5
; Sequence 5, Application US/08846234
; Patent No. 6166292
; GENERAL INFORMATION:
; APPLICANT: OSUMI Chieko
; APPLICANT: KIDA Takao
; TITLE OF INVENTION: RAFFINOSE SYNTHASE GENE, METHOD FOR
; TITLE OR INVENTION: PRODUCING RAFFINOSE, AND TRANSGENIC PLANT
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSGADT, P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,234
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)-413-3000
; TELEFAX: (703)-413-2220
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 784 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-846-234-5

Query Match 100.0%; Score 4225; DB 4; Length 784;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 784; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPSFKNNGSNVVSFDGLNDSSPFAIDGSDFTVNGHSFLSDVPENIVASPSPTYSIDKS 60
DB 1 MAPSFKNNGSNVVSFDGLNDSSPFAIDGSDFTVNGHSFLSDVPENIVASPSPTYSIDKS 60
QY 61 PVSVCGFVFEDASFPDSRNVVSIKGLKDIRFMSIFRFKVMWTHHWGRNGDLESETOY 120
```

```
|||||
Db 61 PVSVCVGDASPDSDRHVYSIGKLDIRFMSIFRFKVMWTTTHWVGRNGDLESETQIV 120
QY 121 ILEKSDSGRPVFLPVEGEFRTSIQGDDEFDVDCVESGSSVYVVASPFSMLYHAGD 180
Db 121 ILEKSDSGRPVFLPVEGEFRTSIQGDDEFDVDCVESGSSVYVVASPFSMLYHAGD 180
QY 181 DPEALVEAMKIVTHTLGTFRLEEKTPPGIVDKFGMCTWDAFYLVHPQGVIEGVRHLV 240
Db 181 DPEALVEAMKIVTHTLGTFRLEEKTPPGIVDKFGMCTWDAFYLVHPQGVIEGVRHLV 240
QY 241 DGCGRPGVLVLDGWSQISGHSDEPTKEGMMQYVAGGOMPRLKFGQENTYFRDYVMPKA 300
Db 241 DGCGRPGVLVLDGWSQISGHSDEPTKEGMMQYVAGGOMPRLKFGQENTYFRDYVMPKA 300
QY 301 TGPRAGOKGMKAFIDELKGEFTEHYVWMLCGYWGGLRPQYVGLPEARVIOPLVSPG 360
Db 301 TGPRAGOKGMKAFIDELKGEFTEHYVWMLCGYWGGLRPQYVGLPEARVIOPLVSPG 360
QY 361 LQMTMEDLAVDKIVLHKVGLVPPKAEEMTEGLHAHLEKVGIDGVKIDVILHLEMLCEDY 420
Db 361 LQMTMEDLAVDKIVLHKVGLVPPKAEEMTEGLHAHLEKVGIDGVKIDVILHLEMLCEDY 420
QY 421 GGRVDLAKAYKAMTKSINKHFKGNGVYASMEHCNDPFLGTBAISLGRVDDFWCTDPS 480
Db 421 GGRVDLAKAYKAMTKSINKHFKGNGVYASMEHCNDPFLGTBAISLGRVDDFWCTDPS 480
QY 481 GDPNGTFMLOGCHVHCANDSLMGNFTHPMDMFQSTHPCAAFAHSAISGSPRIYSD 540
Db 481 GDPNGTFMLOGCHVHCANDSLMGNFTHPMDMFQSTHPCAAFAHSAISGSPRIYSD 540
QY 541 SVGRHNEDLKKLVLPDGSILRSEYALPTRDCLFEDPLHNGETMLKIWNLTKFTGVIGA 600
Db 541 SVGRHNEDLKKLVLPDGSILRSEYALPTRDCLFEDPLHNGETMLKIWNLTKFTGVIGA 600
QY 601 FNCGGGMCRTTRNQCOSQSKRYTSKTNPKDIEHSGENPISIEGKTRALVLYQAKK 660
Db 601 FNCGGGMCRTTRNQCOSQSKRYTSKTNPKDIEHSGENPISIEGKTRALVLYQAKK 660
QY 661 LILSKPSODLIDLPFEFELITVSPVKLIQTSLSHFAPIGILVNLNLSGAIQSVYDQDD 720
Db 661 LILSKPSODLIDLPFEFELITVSPVKLIQTSLSHFAPIGILVNLNLSGAIQSVYDQDD 720
QY 721 LSSVEIGVKGCGEMKRVFASKPRACRIDGDEVGKRYDQDMVVVYVPPIDSSSGISVI 780
Db 721 LSSVEIGVKGCGEMKRVFASKPRACRIDGDEVGKRYDQDMVVVYVPPIDSSSGISVI 780
QY 781 EYLF 784
Db 781 EYLF 784

RESULT 2
US-08-846-234-1
; Sequence 1, Application US/08846234
; Patent No. 6166292
; GENERAL INFORMATION:
; APPLICANT: OSUMI Chieko
; APPLICANT: NOZAKI Jinshi
; APPLICANT: KIDA Takao
; TITLE OF INVENTION: RAFFINOSE SYNTHASE GENE, METHOD FOR
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESS: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT, P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/846,234
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: NORMAN F. OBLON
REGISTRATION NUMBER: 24,618
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)-413-3000
TELEFAX: (703)-413-2220
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 30 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
US-08-846-234-1

Query Match 4.3%; Score 181; DB 4; Length 30;
Best Local Similarity 100.0%; Pred. No. 3e-12;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 215 FGWCTWDAFYLTVPQVIEGVRLVDDGC 244
Db 1 FGWCTWDAFYLTVPQVIEGVRLVDDGC 30

RESULT 3
US-08-770-544-2
; Sequence 2, Application US/08770544
; Patent No. 5907085
; GENERAL INFORMATION:
; APPLICANT: Gonsalves, Dennis
; APPLICANT: Ling, Kai-Shu
; TITLE OF INVENTION: GRAPEVINE LEAFROLL VIRUS PROTEINS AND
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESS: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/770,544
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60009008
; FILING DATE: 21-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1390 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
```


QY 684 ---VSPYTKLIQTSLHAPRIGLV---NMLNTSGAIOGVYDDDLSS-VEIGVKGCGEMRPF 737
DB 920 TLISASEGVSCQNYSEFPSPGJLRESFLDNKPRISGCELYLTWSGLIORHKDSFGHNHY 979
QY 738 ---ASKPRACRHDGEDVGFKYDQDMVVPWPIIDSSSGISVLEY 782
DB 980 SYDAQGLRVKTEODAOYATFEYDNGVRLITTTTKPTTSLSLVTKLEY 1027

RESULT 5
US-08-040-751-3
Sequence 3, Application US/08040751
Patent No. 5407825
GENERAL INFORMATION:
APPLICANT: Payne, Jewel M
TITLE OF INVENTION: Sick, August J
TITLE OF INVENTION: No. 5407825el Bacillus thuringiensis isolates
TITLE OF INVENTION: active against lepidopteran pests and genes encoding No. 54078
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESSES:
ADDRESSEE: DAVID R. SALIWANCHIK
STREET: 2421 N.W. 41st STREET, SUITE A-1
CITY: GAINESVILLE
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/040,751
FILING DATE: 19930329
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SALIWANCHIK, DAVID R.
REFERENCE/DOCKET NUMBER: MA39.C1.D3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1174 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Bacillus thuringiensis
STRAIN: aizawai
INDIVIDUAL ISOLATE: PS81A2
IMMEDIATE SOURCE:
LIBRARY: lambdagem - 11 (tm) Library of August Sick
CLONE: 81A2
US-08-040-751-3

Query Match 2.5%; Score 107; DB 1; Length 1174;
Best local Similarity 19.6%; Pred. No. 0.14;
Matches 111; Conservative 82; Mismatches 181; Indels 192; Gaps 29;

QY 336 YWGGLR-----POVPGIPE-----ARYIQVLSFGIOMTMEDIA-----VDK 372
DB 314 YWAGHRYTSHFTSSQYISSPOYGITANAPSRPIASSTFGIINTLFRILSDPFRSDN 373
QY 373 I-----YAHXGVLVPRKAEAEY--EGLAHNLEKYGIDGVK--IDVILHLEMLCDYDG 421
DB 374 IMPTLGINVGVGFIQPNNGEVLVYRRGTVDSDLELPIIDGENSLVGYSHRL----- 425

QY 422 GRVDLAKAYYKAMFKSI-----NKHEKNGVJASMECHNDF-MPIGTFAI-SLG 468
DB 426 SHVTLRSLNTNITSLPFWTWHHSATDRNIITYPDVITQIPLVKSFSLSLGSVYRGRC 485
QY 469 RVGDDFMCTDPDSCD-----PNCGFMLOGCHM 494
DB 486 FTGGDIIIRTVNNGNVLSMSLNFSNTSLQRYRVRVRYAASQTMVNRVNVGSGTTFDGGFPS 545
QY 495 VHCANDSLWKNFHPDMDFOSTHPCAFHAAISRAISGGPIYVSQVKHNDLLKLIV 554
DB 546 TMSANGSLTQSFRFAEPFGIST-----SGSOTAG-ISISNPGKQTHLRIEF 595
QY 555 LPPGSLIRSEY---VALPFDCLFEEDPLHNGETMLKI---NMLNKEFGYICAFNCOGCG 607
DB 596 IPVDATFEAEYDLERAKAVNSLFTS---SNQLEKTDVTDYHIDOVSNLY---DCLSD 649
QY 608 WCFETRRNQCFS-QYSKRVYSKTN-----PRDIEWHSQ-----EN 641
DB 650 FCUDERKRELSEKKAHKLRLSDERNLQDPNFRGINQPDGWRGSDTDITOGDDVFKEN 709
QY 642 PISIEGV--KTFALYLYQAKKILSK-----PSODDIALDPF--EFLIT 683
DB 710 YVTLPGTFDECYPTLYXO--KIDSKLKAYNRYQLRGYIEDSODLEIYLYRYNAKHETVN 767
QY 684 VSPYTKLIQTSLHAPRIGLVNMLNTSGAIOGVYDDDLSS-VEIGV 727
DB 768 VPGTGSIMPLSVE-SPIGRCGEPRN--CVPHELMNPDLDCSCRDGEKCAHSHHSFLID 824

QY 728 VKGCGEMRVFASKPRACRIDGEDVG 753
DB 825 V-CTTDLQ-----EDLG 835

RESULT 6
US-08-291-368-2
Sequence 2, Application US/08291368
Patent No. 5686069
GENERAL INFORMATION:
APPLICANT: Payne, Jewel M.
TITLE OF INVENTION: No. 5686069el Bacillus thuringiensis isolates
TITLE OF INVENTION: Active Against Lepidopteran Pests
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESSES:
ADDRESSEE: SalIWanchik & SalIWanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: US
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/291,368
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION NUMBER:
APPLICATION NUMBER: 07/597,607
FILING DATE: 15-OCT-90
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: SalIWanchik, David R.
REGISTRATION NUMBER: 31,794
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904)375-8100
TELEFAX: (904)372-5800
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:


```

; LENGTH: 1174 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Bacillus thuringiensis
; STRAIN: aizawai
; INDIVIDUAL ISOLATE: PS81A2
; LIBRARY: Lambdaagem - 11 (tm) Library of August Sick
; CLONE: 81A2
US-08-291-368-2

```

```

Query Match 2.5%; Score 107; DB 1; Length 1174;
Best local Similarity 19.6%; Pred. No. 0.14;
Matches 111; Conservative 82; Mismatches 181; Indels 192; Gaps 29;

```

```

QY 336 YMGGLR-----POVPLPE-----ARVQPVLSPGLOMTMEDA-----VDK 372
DB 314 YMAGHRYTSHFTGSSQVYSSPOYGITANAEPSRTIAPSTFGINLEVRTLSDEFFRRSDN 373
QY 373 I-----VHKVGLVPEKAEEM--EGLHALEKVGIDGVK--IDVHILMLCEDYG 421
DB 374 IMPTLGINVOGVGFLOPNNGEVLVRRRGVDSLDLPIDGENSLVGYSHRL----- 425
QY 422 GRVDLAKAYKAMTKSI-----NKHFGNGVIASMEHCNDF--MFLGTEAI--SLG 468
DB 426 SHVTLRSLVNTNITSLPTFVWTHHSATDRNIITPDVITQIPLVKSPSLSGTSVVRGPG 485
QY 469 RVDDDFWCIDPSCD-----PMTWLOGCHM 494
DB 486 FTGCDIIRTVNNGVLSMSLNFSNTSLQRYRVRYAASOTMVRVAVGSGTTFDQGFPS 545
QY 495 VHCANDSLMNGNFIHPDMFQSTHPCAFHNASRAISGPIVVSQVSGKHNFDLKLKV 554
DB 546 TMSANGSLTSQSFRFAEPVGIS-----SGSOTAG--ISISNPNQGTJHLDRIEF 595
QY 555 LPDGSILRSEY---VALPTDCLEFEDPLHNGETMLKI---WNLNKTVGIAGFNCQGG 607
DB 596 IPVDATFEAYEDLERAKAVNSLFTS---SNOJELKTDVTDYHIDQVSNLY---DCLSD 649
QY 608 WCEETRANOCFS-QYSKRVTSKTN-----PKDIENHSG-----EN 641
DB 650 FCLDERKELSEKVKHAKRLSDERNLQDPNFRGINRQPRGMRGSTDITIOGDDVFEKN 709
QY 642 PISIEGV--KTFALYLOAKKILSK-----PSODLDIALDPF--EPELIT 683
DB 710 YVLPFGFDECPTTYLQ--KIDESKLIKAYNRQOLKGIYDSQDLLEYLIRNAKHETTV 767
QY 684 VSBVTKLQTSLHFAPIGLVNMMLNTSGAIVSDYDDLS-----SVEIG 727
DB 768 VPGGSLMPLSVE--SPIGRGEPNR--CVRPHLEWNPDLDCSDBGEKCAHSHHFLSDID 824
QY 728 VKGCGEAKRVASKKPRACRIDGCVG 753
DB 825 V-GCTDLDQ-----EDLG 835

```

```

RESULT 7
US-08-962-190-2
; Sequence 2, Application US/08962190
; Patent No. 5985267
; GENERAL INFORMATION:
; APPLICANT: Payne, Jewel M.
; APPLICANT: Sick, August J.
; TITLE OF INVENTION: No. 5985267el Bacillus thuringiensis Isolates
; TITLE OF INVENTION: Active Against Lepidopteran Pests
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: US
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/962,190
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/291,368
; FILING DATE:
; APPLICATION NUMBER: 07/597,607
; FILING DATE: 15-OCT-90
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: MA50.C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904)375-8100
; TELEFAX: (904)372-5800
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1174 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Bacillus thuringiensis
; STRAIN: aizawai
; INDIVIDUAL ISOLATE: PS81A2
; LIBRARY: Lambdaagem - 11 (tm) Library of August Sick
; CLONE: 81A2
US-08-962-190-2

```

```

Query Match 2.5%; Score 107; DB 2; Length 1174;
Best local Similarity 19.6%; Pred. No. 0.14;
Matches 111; Conservative 82; Mismatches 181; Indels 192; Gaps 29;

```

```

QY 336 YMGGLR-----POVPLPE-----ARVQPVLSPGLOMTMEDA-----VDK 372
DB 314 YMAGHRYTSHFTGSSQVYSSPOYGITANAEPSRTIAPSTFGINLEVRTLSDEFFRRSDN 373
QY 373 I-----VHKVGLVPEKAEEM--EGLHALEKVGIDGVK--IDVHILMLCEDYG 421
DB 374 IMPTLGINVOGVGFLOPNNGEVLVRRRGVDSLDLPIDGENSLVGYSHRL----- 425
QY 422 GRVDLAKAYKAMTKSI-----NKHFGNGVIASMEHCNDF--MFLGTEAI--SLG 468
DB 426 SHVTLRSLVNTNITSLPTFVWTHHSATDRNIITPDVITQIPLVKSPSLSGTSVVRGPG 485
QY 469 RVDDDFWCIDPSCD-----PMTWLOGCHM 494
DB 486 FTGCDIIRTVNNGVLSMSLNFSNTSLQRYRVRYAASOTMVRVAVGSGTTFDQGFPS 545
QY 495 VHCANDSLMNGNFIHPDMFQSTHPCAFHNASRAISGPIVVSQVSGKHNFDLKLKV 554
DB 546 TMSANGSLTSQSFRFAEPVGIS-----SGSOTAG--ISISNPNQGTJHLDRIEF 595
QY 555 LPDGSILRSEY---VALPTDCLEFEDPLHNGETMLKI---WNLNKTVGIAGFNCQGG 607
DB 596 IPVDATFEAYEDLERAKAVNSLFTS---SNOJELKTDVTDYHIDQVSNLY---DCLSD 649

```

```

0Y 608 WCRETRNNGCES--OYSRRTVSKTN-----KRIEWHSG-----EN 641
Db 650 FCLDEKKELSKAKVHARLSDERULLDDPPNFRGINRQPDWGMGSGSDITITGGDDVYFKEN 709
QY 642 PISIEGV--KTFALTYLYOAKKILSK-----PSODLDIALDPF--EPELIT 683
Db 710 YVTLPGFDECEYPTLYYO--KIDESKLIKAVNRYQLRGYIEDSODLEIYLLRYNAKHETVN 767
QY 684 VSPPTKLIQSLHNPAPLGYVLMNTSGAIGSVYDDLS-----SVEIG 727
Db 768 VPGTGSIMPLSLVE--SPIGRCGEPPNR--CVPHELEWNPDLDCS CRDEKCAHHSHHFSLID 824
QY 728 VKGCGEMRFEASKRPRACRIDGEDVUG 753
Db 825 V-GCTDLO-----EDLG 835

RESULT 8
PCT-US95-10310-2
: Sequence 2, Application PC/WTUS9510310
GENERAL INFORMATION:
APPLICANT: MYCOGEN CORPORATION
APPLICANT: STREET ADDRESS: 5501 Oberlin Drive
APPLICANT: CITY: San Diego
APPLICANT: STATE/PROVINCE: California
APPLICANT: COUNTRY: US
APPLICANT: POSTAL CODE/ZIP: 92121
APPLICANT: PHONE NUMBER: (619) 453-8030
APPLICANT: FAX NUMBER: (619) 453-6991
TITLE OF INVENTION: Protein Toxins Active Against Lepidopteran Pests
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Saliwanchik & Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: US
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10310
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/597,607
FILING DATE: 15-OCT-90
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: MA50.C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904) 375-8100
TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1174 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Bacillus thuringiensis
STRAIN: aizawai
INDIVIDUAL ISOLATE: PS81A2
IMMEDIATE SOURCE:

```

```

;      LIBRARY: LambdaGem - 11 (tm) Library of August Sick
;      CLONE: 81A2
PCT-US95-10310-2

```

Query Match	2.5%	Score 107;	DB 5;	Length 1174;
Best Local Similarity	19.6%	Pred. No. 0.14;		
Matches 111;	Conservative 82;	Mismatches 181;	Indels 192;	Gaps 29;

```

QY 336 YWGLR-----POVGLPE-----ARVIOVLSPGLOMTMEDA-----VK 372
D 314 YWAGHRTSHFTGSSOYISSFOGITAANAPRSRTIASTFTPGILTEYRLTSDFFFRSD 373
QY 373 I-----VLHKVGLVPEKAEMEY-EGJHAHLEKVGIDVK-IDVTHLEMLCEDYG 421
D 374 IMPTLGINVVOGVFIQPNNGEVLXRRGTVDSDLEPLDGENSLVGYSHRL----- 425
QY 422 GRVDLAKVYKAMTKSI-----NKHFGNGVYIASEMHCNPF-MLTGETAI-SLG 468
D 426 SHVLTFRSLXNTNTISLPTFVWTHHSATDRNIIYDPVITQIPLVKSFTSLGTSVVRGP 485
QY 469 RVGDPMCTDPSGD-----PAGTWMLOGCHM 494
D 486 FTGDDITRTNWNGVLSMSLNFSTSLORTRVRVRYAASOTMWRVAVGSGSTFFDGGFPS 545
QY 495 VHCANDSLMGNFTHPDMFQSTHPCAEFHAASRAISGPIYVSDSVGKHNFDLKLTV 554
D 546 TMSANGJLTQSFPFAEFPGVISTP-----SGSQTAG-ITSINNPGKOTFHLDRIEF 595
QY 555 LPDGSILRSEY---YALPTRDCLFEDPLHNGETMLK---WMNKTKTVIGAFNQGCG 607
D 596 IPVATFEAEEDLERACKAVASLFTS---SNQIELKTDVTDYHNDVSNLY---DCLSD 649
QY 608 WCRETRNROGCS-OYSKRVTSKTN-----PKDIEHNSG-----EN 641
D 650 FCLDEKRELSTKVKHAKRLSERMLDODPNFRGINRQPRDGKWRGSDITITOGGDDVFKEN 709
QY 642 PISIEGV-KTFALYLYOAKMLLK-----PSODLDIADLPF-EFELLIT 683
D 710 YVTLPGFMEDECYPTLYLO-KIDESKLTAKYNRVOLRGYIEDSODELEYLILRYNAKHETVN 767
QY 684 VSPYTKLIQTSLHAPRIGLVMLMTSCAIGSDVYDDLS-----SVEIG 727
D 768 VPGTGSIMPLSVE-SPIGRCGEPRN-CVPHLEWNPDLDCSRDEKCAHSHHFFSLDID 824
QY 728 VKGGEMKRVFASKKPRACRIDGEDVG 753
D 825 V-GCTDLO-----EDLG 835

```

```

RESULT      9
5164180-4
; Patent No. 5164180
; APPLICANT: Payte, Jewel, Sick, August J.
; TITLE OF INVENTION: BACILLUS THURINGIENSIS ISOLATES ACTIVE
; AGAINST LEPIDOPTERAN PESTS
; NUMBER OF SEQUENCES: 6
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/451,389
; FILING DATE: 14-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 353,860
; FILING DATE: 18-MAY-1989
; SEQ ID NO:4:
; LENGTH: 1174
5164180-4

```

Query Match	2.5%	Score 107;	DB 6;	Length 1174;
Best Local Similarity	19.6%;	Pred. No. 0.14;		
Matches 111; Conservative	82;	Mismatches 181;	Indels 192;	Gaps 29;
QY 336 YWGGAR-----PQVPELPE-----ARVIGQPVLSPGIQTMTMEDLA-----YDK 372				

```

DB 314 YNAGHRTVSHFTGSSQVIVSSPOGITANAEPSRTIAPSTFPGILNLEYRLSDPFRSRDN 373
QY 373 I-----VLHRYGLVPPKAEEMV--EGLHAHLEKVGIDGVK--IDVHILHLEMCEDDG 421
DB 374 IMPTLGINVAGGFGTQPNNGEVLVRRGTVDLDELPIGENSELGYSHRL----- 425
QY 422 GRVDLAKAYYKAMTKSI-----NKHFGNGVIAEMHCNDF-MFLGTEAI-SLG 468
DB 426 SHVTLRLSLNTNITSLPFWVTHHSATDRNITYPVITQIPLYKSFSLSGISVVRGFG 485
QY 469 RVGDDEFWCTDPSGD-----PNGTFWLOGCHM 494
DB 486 FTGGDIIRTNVNGNVLMSLNFSTSLQRYRVRYVRYAASQTMVRYVNGSGSTFGDGFPS 545
QY 495 VHCANSLMKNFNIHPDMFMFQSTHPCAFHMAASRAISGPIYVSQVGNHNDLAKLV 554
DB 546 TMSANGSLTSSQSFRAEFVPGIST-----SGSQTAG-ISISNPNRQTFHLDRIF 595
QY 555 LPDGSILRSEY---VALPTRDCLFEDPLHNGETMLKI-----NMLNKFQVIGAFNCOGG 607
DB 596 IYPDAIFEFAYDIERAKAVNSLFTS---SNQLEKTDVTHIIDVSNLV---DCLSD 649
QY 608 WCRFTRNOCFS-QYSKRYTSKTN-----PRDIEMHSG-----EN 641
DB 650 FCLDEKRELSEKKAHAKRLSDERNLLQDPNFRGINRQPRGWRGSDITIGQDDVFKEN 709
QY 642 PTISIEGV--KTALVLYQAKKILSK-----PSODDIALDPF--EPELIT 683
DB 710 VYVLPCTFDECFTYLYQ--KIDESKLKAYNRQYQLGYIEDSDLEIYLRVNAKHETVA 767
QY 684 VSPVTKLIOTSLHFAPIGLVNLMTSGAIOQSVYDDLS-----SVEIG 727
DB 768 VEGTGLMPLSVE-SPTRGCGEPNR--CVPHLEWNPDLQSCDGEKCAHSHHESLDID 824
QY 728 VNGCGEMRVASKKPRACRIDGEDVG 753
DB 825 V-GCTDLQ-----EDLG 835

RESULT 10
: Sequence 2, Application US/08846234
: Patent No. 6166292
: GENERAL INFORMATION:
: APPLICANT: OSUMI Chieko
: APPLICANT: NOZAKI Jinshi
: APPLICANT: KIDA Takeo
: TITLE OF INVENTION: RAFFINOSE SYNTHASE GENE, METHOD FOR
: TITLE OF INVENTION: PRODUCING RAFFINOSE, AND TRANSGENIC PLANT
: NUMBER OF SEQUENCES: 22
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT, P. C.
: STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
: CITY: ARLINGTON
: STATE: VIRGINIA
: COUNTRY: USA
: ZIP: 22202
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/846,234
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: NORMAN F. OBLON
: REGISTRATION NUMBER: 24,618
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703)-413-3000
: TELEFAX: (703)-413-2220
: INFORMATION FOR SEQ ID NO: 2:

```

```

: SEQUENCE CHARACTERISTICS:
: LENGTH: 19 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: FRAGMENT TYPE: internal
US-08-846-234-2

```

```

Query Match 2.5%; Score 105; DB 4; Length 19;
Best Local Similarity 100.0%; Fred. No. 0.00014;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 61 PVSVCGFVGFDASEPDSRH 79
DB 1 PVSVCGFVGFDASEPDSRH 19

```

```

RESULT 11
US-08-305-505-6
: Sequence 6, Application US/08305505
: Patent No. 5668001
: GENERAL INFORMATION:
: APPLICANT: Miziorko, Henry M.
: TITLE OF INVENTION: 3-HYDROXY-3-METHYLGUTARYL-COA
: TITLE OF INVENTION: SYNTHASE PREPARATION WITH IMPROVED
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Quarles & Brady
: STREET: 411 East Wisconsin Avenue
: CITY: Milwaukee
: STATE: Wisconsin
: COUNTRY: U.S.A.
: ZIP: 53202
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/305,505
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/072,040
: FILING DATE: 02 JUNE 1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Baker, Jean C.
: REGISTRATION NUMBER: 35,433
: REFERENCE/DOCKET NUMBER: 65-053-9083-9
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (414) 277-5709
: TELEFAX: (414) 277-5591
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 520 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-305-505-6

Query Match 2.4%; Score 100; DB 1; Length 520;
Best Local Similarity 19.0%; Fred. No. 0.18;
Matches 82; Conservative 59; Mismatches 111; Indels 180; Gaps 20;
QY 378 VGLVPE-----KAE-EMTEGLHAHLEKVGIDGVKID-----VIHL 412
DB 17 VGIVALEIYFPQYVDQALEKYGVDAGRYTIGLQAKKGFCTDRDINSICMTVQNL 76
QY 413 LEM--LCEDYGRVDLAKAYYKAMTKSINKH-----KNGVIAEMHCNDFMFLGTEA 464

```

```

Db      77 MERNNLSDYDCIGRLEVGTEITIDKSKSVKTNLMOLFEEESGNTDIEGIDTNN-ACYGGTAA 135
QY      465 ISLRGVGDFFNCTPDSGPNCTFPMLOGCHMVHCANDSLMGNFTHPDMQOSTHPCAA 524
Db      136 VF-----NAVNWIE-----SSWDGRT-----ALVAGDIAVY 163
QY      525 HAASRAISGPIYVSDVG-----KHNFDLLKLVLPDGSILRSEYAL 568
Db      164 ATGNARPTGGVAVALLIGPNAPLIFERGLRGTHMCHAYDYK---PD---MLSEY--- 213
QY      569 PTRCLFEDPLHNGETMLKINLNKFTVIGAFNCGGCKRETRNCCFSGYSKRYTSK 628
Db      214 -----PVDGKLSIQ-----CYLSALDRCYSYCKKHA- 242
QY      629 TNPKDIEHSENPISIGVTFALYQAKKLLSKPSODLDIALDPFEELIVSPVT 688
Db      243 -----QW-----QKFNANDPFLNDPFGF-MIHSFYC 268
QY      689 KLIOGSLHFAPIGLVNMNTSGAIOVDYDDDLSSVEIGVGCGEMRVFASKKPRACRID 748
Db      269 KLVQKSL--AMMLNDPFLN-----DQNRDKNSIYSGLAFGDVKL----- 306
QY      749 GEDVGFKYDQDQ 760
Db      307 -EDTYFDRDVER 317

```

```

RESULT 12
US-08-717-515-8
; Sequence 8, Application US/08717515
; Patent No. 5736373
; GENERAL INFORMATION:
; APPLICANT: Hamilton, Paul T.
; TITLE OF INVENTION: THERMOSTABLE DNA POLYMERASE FROM
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: R. J. Rodrick, Becton Dickinson and Company
; STREET: 1 Becton Drive
; CITY: Franklin Lakes
; STATE: NJ
; COUNTRY: US
; ZIP: 07417
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/717,515
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fugitt, Donna R.
; REGISTRATION NUMBER: 32,135
; REFERENCE/DOCKET NUMBER: P-3544
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1276 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-717-515-8

```

```

Query Match 2.4%; Score 100; DB 1; Length 1276;
Best Local Similarity 21.2%; Pred. No. 0.87;
Matches 116; Conservative 74; Mismatches 192; Indels 166; Gaps 31;
QY 311 KAFIDELGGEFTVYHVVHVALCGYMGRLRQVPLGEA--RVIOPLVSLQLOMTMEDL 368
Db 84 KAFODKL-----YPTWDAY--RYNGKLIAYIAVEALSLYKDLPLNPPTWEEI 133

```

```

QY      369 -AVDKIVLHKVGLVPPERKEEMV-----EGCHA-----HLEKVGTD--GVKTDV 409
Db      134 PALDELKAKKKSALMFNLDPYFTWPLIADGYAFAFYENGKYDIKDYVDNAGAKGGL 193
QY      410 IHLEMLCE-----DYG-----GRVDAKAYYKAMTKSINKH 441
Db      194 TELVDLKNKHNADTDVSIAEAFNKGETAMTINGPWAMSNIDTSKVNYV---TVLPT 250
QY      442 FKG-----NGVIASMEHCNDFM--FL---GTEAIS---LGRVDDFWCT 477
Db      251 FKQPSKRFVGLVSGIAAASPKNELAKEFLBNLYLTDEGLEAVNKKPLCAVALKSYEE 310
QY      478 DPSGDPNGTFMLQGCCHVHCANDSLMGNFTHPDMQFQSTHPCAAFLAAS-----RAI 531
Db      311 ELAKDPRIAAIMEN-----AOKGEIMPNI--PQMSAFWYAVRTAVIYAASGRQTVDEAL 362
QY      532 SGGPIYVSDSYGKH-----FDLLKLVLPDG--SILRSEYALPTRD---CL 574
Db      363 KDAQTNSSNNNNNNNNNNLIEGRISFEGVTKLVLDGNSIAYRAFFALPLLNNDKGI 422
QY      575 FEDPLHNGETM-LKIMNLKFTGVIGAFNCGGWCRETRNCCFSGYSKRYTSKTNPKD 633
Db      423 YTNALYGTNMLKVLBEERKPHILVAFDAG---KTFRHKTFKEV-KGRQKXTPE- 475
QY      634 IEWHSGENP-----ISIEGVTFALYQAKKL---LSKPSODLDIA-----LDPF 677
Db      476 --LSEQLPFIKDLIDAVQIYTYELENEYADDIIGTVARQAEKDOFEVKIISGKDLQOL 532
QY      678 EFELTVSPVTK-----LIOTSLHFAPIGLVNMNTSGAIOVDYDDDLSSVE 725
Db      533 ATEKTTVS-ITKKGITDVEPHPTPESTQEKYGLSPAQIIDLGLMG-----DQSDNIP--- 583
QY      726 IGKVGCGE 733
Db      584 -GVPGVGE 590

```

```

RESULT 13
US-08-525-654A-1
; Sequence 1, Application US/08525654A
; Patent No. 5736356
; GENERAL INFORMATION:
; APPLICANT: SANO, KOICHIRO
; APPLICANT: KUMAZAWA, YOSHIIYUKI
; APPLICANT: YASEBUDA, HISASHI
; APPLICANT: SEGURO, KATSUYA
; APPLICANT: MOTOKI, MASAO
; TITLE OF INVENTION: TRANSGUTAMINASE ORIGINATED FROM
; NUMBER OF SEQUENCES: 150
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/525,654A
; FILING DATE: 28-SEP-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 6/8283
; FILING DATE: 28-JAN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7/3876

```

FILING DATE: 13-JAN-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: OBLON, NORMAN F.
 REGISTRATION NUMBER: 24,618
 REFERENCE/DOCKET NUMBER: 10-760-0 PCT
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 703-413-3000
 TELEFAX: 703-413-2220
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 770 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 ORIGINAL SOURCE:
 ORGANISM: Crassostrea gigas
 US-08-525-654A-1

Query Match 2.3%, Score 98.5; DB 1; Length 770;

Best Local Similarity 19.5%, Pred. No. 0.51; Mismatches 241; Indels 239; Gaps 35;

Matches 134; Conservative 72; Mismatches 241; Indels 239; Gaps 35;

261 DSDPTK-EGMNOT-----VAGEOMPRLKFOEN-----YKFRDYVPKATGP-----303
 147 DLPPTKSDVETETFAEYKNAAGKLECAITSAPOSSIGETRYETNLKTDVAKRYE 206
 304 -----RAGOKMKAFIDE--LKGEFTEHYVYMHALCGYGLRPQVGLPEARV 352
 207 ENESMILIFNMAKEDTYMDEKGERGEVLENGRVTSTRT--WYG-RPMNFG-----QF 259
 353 IQPVLSPLQMTMEDLAVDKYLHVGLVPPKAEEMTEGLHAHLEKVGIDGVKIDVYHL 412
 260 DDPVLDVALQLLEGG-----GLSDVACTSP-----VSVIRC 290
 413 LEMCEDYGGRRDLAKAYYKATKSIKHFKNGVYASNE--HCND-----FMFLG 461
 291 LSSLCNSCDNMGVLAGRTKRYKPCQYPMKWTGSAVLIKEHTGNKPRYRGQCMVSG 350
 462 TEAISLGRVG-----DDEFCTD--PSGDPNGTFWLOGCHMVCA 498
 351 LITWCRCLGIPTRSVTFNEDSAHDTSSMTIDSHWDEDEPLEDNDSVW-----NFHW 405
 499 NDSLAMGNFIHPD--WDMFOSTHPCAFHNASR-----AISGPIYSDSV-----G 543
 406 NES-WFRRLDPEGYDGOAHADAT--POEASEGIMRCGAPRLAIKEGHVYLNFDIPVFG 463
 544 KHNFDLKKLVLPDQSI-----DDEFCTD--PSGDPNGTFWLOGCHMVCA 560
 464 EYNGDRVQWYKKGDTMEVSOIDSAVGHYSTKPMGSDNDRDVTNLYKYPGSGEORRV 523
 561 -----LSEYVALPTRDCLF-----EDPLHNGETMLKINLNKFTG-----VI 598
 524 AKFYVRYSYTRRQNTIKYLDTRKVLKFTVSPDNTLIGDMEIKVAVKMTADPLKLHLTV 583
 599 GAFNCGGGMCHETRRNCFQSKRYTSK--TNP-KDIEHSGENPISIEGVKFPALYL 655
 584 SLVNYVYGVAGSRVKTQTFEETIKAKDEKIVTPVKTEYHDGNP--EG--RFOLYI 638
 656 YQAKKL-----ILSKPSQDLIALDPFEF-----LITVSPYKLIQTSIH 696
 639 -SGKNIDSGMESIOISVLPKPOLVIOV--PQTEAKETETATVKNTOVLVTOAE 694
 697 FAPLIVLMLNTSG--AIQSYVDYDDLSVEIGVGCGERVFAASKPRACRIDEDYGF 754
 695 IAVES-----SGLLAQOTID-----ISSPIKPGDEVKTYVLRKRPYIMGRBLIA 740
 755 KYDDQDMVYVQVPMPIIDSSSGGISVI 780
 741 TETSKOIVDIE-----TSADIKVI 759

RESULT 14

US-08-525-654A-3

; Sequence 3, Application US/08525654A

; Patent No. 5736356

; GENERAL INFORMATION:

; APPLICANT: SAMO, KOHICHIRO

; APPLICANT: KUMAZAWA, YOSHIYUKI

; APPLICANT: YASEUDA, HISASHI

; APPLICANT: SEGURA, KATSUYA

; APPLICANT: MOTOKI, MASAO

; TITLE OF INVENTION: TRANSGLUAMINASE ORIGINATED FROM

; NUMBER OF SEQUENCES: 150

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,

; ADDRESS: P.C.

; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400

; CITY: ARLINGTON

; STATE: VA

; COUNTRY: USA

; ZIP: 22202

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent Release #1.0, Version #1.30

; CURRENT APPLICATION DATA: US/08/525,654A

; FILING DATE: 28-SEP-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 6/8283

; FILING DATE: 28-JAN-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 7/3876

; FILING DATE: 13-JAN-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: OBLON, NORMAN F.

; REGISTRATION NUMBER: 24,618

; REFERENCE/DOCKET NUMBER: 10-760-0 PCT

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 703-413-3000

; TELEFAX: 703-413-2220

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 771 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; ORIGINAL SOURCE:

; ORGANISM: Crassostrea gigas

; US-08-525-654A-3

Query Match 2.3%, Score 98.5; DB 1; Length 771;

Best Local Similarity 19.5%, Pred. No. 0.51; Mismatches 241; Indels 239; Gaps 35;

Matches 134; Conservative 72; Mismatches 241; Indels 239; Gaps 35;

261 DSDPTK-EGMNOT-----VAGEOMPRLKFOEN-----YKFRDYVPKATGP-----303
 148 DLPPTKSDVETETFAEYKNAAGKLECAITSAPOSSIGETRYETNLKTDVAKRYE 207
 304 -----RAGOKMKAFIDE--LKGEFTEHYVYMHALCGYGLRPQVGLPEARV 352
 208 ENESMILIFNMAKEDTYMDEKGERGEVLENGRVTSTRT--WYG-RPMNFG-----QF 260
 353 IQPVLSPLQMTMEDLAVDKYLHVGLVPPKAEEMTEGLHAHLEKVGIDGVKIDVYHL 412
 261 DDPVLDVALQLLEGG-----GLSDVACTSP-----VSVIRC 291
 413 LEMCEDYGGRRDLAKAYYKATKSIKHFKNGVYASNE--HCND-----FMFLG 461

```

Db 292 LSSLCNSCDNNGVLAGKMTKEPRDCTVPKMTGSAVIAIKEYHTNGKPVRYGQCWVFSG 351
QY 462 TEALISGRVG-----DDFWCTD--PSGDPNGCTFWLOGCHMVHCA 498
Db 352 LITTRCRCLGIPRSTVNFPSAHDDBSSMTIDSHMDGDGEPLDMNSVM-----NHHW 406
QY 499 NDSLMMGNFIHPD--WDMFOSTHPCAFHAASR-----AISGPIYVDSV---G 543
Db 407 NES-WFRRLDLPFGYDGMQAHDAF--PQASEGIMRCGPAPLTAIKESHVYLNFDIPFVFG 464
QY 544 KHPDILKIKVLDGSI-----560
Db 465 EVNGDRVQWVYKDGMEVSQIDHSAYGHVISTKRMGSDNREDVTNLKYPDGSQERRV 524
QY 561 -----LRSEYALPFRDCLF-----EDPLHNGETMLKIMNKKFTG-----VI 598
Db 525 AKVYNRSTRKQNTYLTDTTKYKFTVSPDNLIDDMEIKAVANTADPLKLLTV 584
QY 599 GAFCNGCGGWCRTRRNOCFSQYKRVTSK--TNP-KDIEMHSGENDISIEGVKTFALYL 655
Db 585 SLVNAYYTVAGSRVKQTPEETIKAKDEKIVMPVKGTGYHDGMP---EG--RFQLYI 639
QY 656 YQAKKL-----ILSKPSQDDIALDPREF-----LITVSPYTKLIQTSLSH 696
Db 640 -SGKNIDSGSESTQISFVLKPKQLVLYQ---PQTIKAKEETEAIVFKNTQVLVLYQAE 695
QY 697 FAPIGLYNMLNTSG--AIGSVYDDDLSSVEIGVKGCGEMKRVFASKPRACRIDGEDVGF 754
Db 696 IAYEG-----SGLAPQID-----ISSPIKPDDEVKKTIVLRKRPYYGRRELLIA 741
QY 755 KYDQDQVVVQVWPVIDSSSGGISVI 780
Db 742 TETSKQIVDIE-----TSADIKVI 760

RESULT 15
US-08-305-505-4
; Sequence 4, Application US/08305505
; Patent No. 5668001
; GENERAL INFORMATION:
; APPLICANT: Mizoroku, Henry M.
; TITLE OF INVENTION: 3-HYDROXY-3-METHYLGITARYL-COA
; TITLE OF INVENTION: SYNTHASE PREPARATION WITH IMPROVED
; TITLE OF INVENTION: STABILITY
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: Wisconsin
; COUNTRY: U.S.A.
; ZIP: 53202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/305,505
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/072,040
; FILING DATE: 02 JUNE 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Baker, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 65-053-9083-9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5709
; TELEFAX: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 520 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-305-505-4

Query Match 2.3%; Score 97; DB 1; Length 520;
Best Local Similarity 18.8%; Pred. No. 0.36;
Matches 80; Conservative 63; Mismatches 115; Indels 168; Gaps 21;

QY 378 VGLVPE-----KAE-EMTEGLHAHLEKVGIDGVKD-----YIHL 412
Db 17 VGIYALEIFPSQYVDDAELEKXYDVADGKYYTIGLGOARMGFCTDRDINSCLTVVQNL 76
QY 413 LE-MLCEDYGGRYDLAKAYKAMTKSINKHF-----KNGCVIASMEHCNDFEFLGTEA 464
Db 77 MERNSLSYDCIGRLVGTETITIDKSKSVKSNLMOLFESGNTDIEGIDTTN-ACYGGTAA 135
QY 465 ISLGRVGDDEFWCTDPSGDPNGTFWLOGCHMVHCAANDSLMMGNF---IHPWDMFQ--STH 519
Db 136 VF-----NAVNIE-----SSWDGRVALVAGDAIAYATGNAR 169
QY 520 PCAAFHAASRAISGPIYVSD-----SYGKHNFDLKLVLPDGSILRSEYALPTRDCL 574
Db 170 PTGGVGAVALLIGNPALPIEDRGLRGTHMOHAYDFYK---PD--MLSEY-----213
QY 575 FEDPLHNGETMLKIMNKKFTGVIGAFNCGGWCRTRRNOCFSQYKRVTSKTNPKDI 634
Db 214 ---PIVDKLSIQ-----CYISALDRCSYVRKKIRA-----242
QY 635 EMHSGENDISIEGVKTFALYLYQAKKLILSKPSQDDIALDPFEFELITVSPYTKLIQTS 694
Db 243 QW-----QREGNDNFTLNDFGF-MISHSPYCKLYOKS 274
QY 695 LHFAPIGLYNMLNTSGAISOVDYDDDLSSVEIGVKGCGEMKRVFASKPRACRIDGEDVGF 754
Db 275 L-ARMLNDPLN-----DQNRKNSIYSGLEARGDVKL-----EDTYF 311
QY 755 KYDQDQ 760
Db 312 DRDYER 317

```

Search completed: November 30, 2001, 09:43:45
 Job time: 204 sec

c

